PHARMACOMETRICS

The Science of Quantitative Pharmacology

Edited by: **Ene I. Ette & Paul J. Williams**

PHARMACOMETRICS

THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

*E*ach generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

PHARMACOMETRICS
THE SCIENCE OF QUANTITATIVE PHARMACOLOGY

Edited by

Ene I. Ette
Anoixis Corporation

Paul J. Williams
University of the Pacific and Anoixis Corporation

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at 877-762-2974, outside the United States at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my wife, Esther, who supports, comforts, and inspires and is always there for me.
E. I. E.

To my wife, Debbie, who supports, comforts, and inspires.
P. J. W.

CONTENTS

Alaa Ahmad, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 [alaa_ahmad@vrtx.com]

Peter L. Bonate, Genzyme Corporation, Pharmacokinetics, 4545 Horizon Hill Blvd., San Antonio, TX 78229 [peter.bonate@genzyme.com]

Daniel Brazeau, Department of Pharmaceutical Sciences, 517 Cooke Hall, State University of New York at Buffalo, Buffalo, NY 14260 [dbrazeau@buffalo.edu]

René Bruno, Pharsight Corporation, 84 Chemin des Grives, 13013 Marseille, France [rbruno@pharsight.com]

Edmund V. Capparelli, Pediatric Pharmacology Research Unit, School of Medicine, University of California—San Diego, 4094 4th Avenue, San Diego, CA 92103 *and* Trials by Design, 1918 Verdi Ct., Stockton, CA 95207 [ecapparelli@ucsd.edu]

Hui-May Chu, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 [hui-may_chu@vrtx.com]

Valérie Cosson, Clinical Pharmacokinetics Modeling and Simulation, Psychiatry, GSK Spa, Via Fleming 4, 37135 Verona, Italy [valerie.2.cosson@gsk.com] *and* Hoffman—La Roche Ltd., PDMP, 663/2130, CH-4070 Basel, Switzerland [valerie.cosson@roche.com]

Charles W. Dement, 260 Jacobs Management Center, University at Buffalo–SUNY, Buffalo, NY 14260

Chantaratsamon Dansirikul, School of Pharmacy, University of Queensland, Brisbane 4072, Australia [joy@pharmacy.uq.edu.au] *and* Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden

Stephen B. Duffull, School of Pharmacy, University of Queensland, Brisbane 4072, Australia [sduffull@pharmacy.uq.edu.au] *and* School of Pharmacy, University of Otago, PO Box 913, Dunedin, New Zealand [stephen.duffull@stonebow. otago.ac.nz]

Ene I. Ette, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 *and* Anoixis Corp., 214 N. Main St., Natick, MA 01760 [ene_ette@anoixiscorp.com]

Wayne Ewy, Pfizer, PGR&D, 2800 Plymouth Road, Ann Arbor, MI 48105 [wayne.ewy@pfizer.com]

Farkad Ezzet, Pharsight Corporation, 87 Lisa Drive, Chatham, NJ 07928

Emmanuel O. Fadiran, Division of Clinical Pharmacology 2, OCP, FDA, 10903 New Hampshire Avenue, Building 21, Silver Springs, MD 20993-0002 [emmanuel.fadiran@.fda.hhs.gov]

Jill Fiedler-Kelly, Cognigen Corporation, 395 S Youngs Rd., Williamsville, NY 14221 [Jill.Fiedler-Kelly@cognigencorp.com]

Bill Frame, C.R.T., 5216 Pratt Rd., Ann Arbor, MI 48103 [framebill@ameritech. net]

Gilles Freyer, Ciblage Thérapeutique en Oncologie, Service Oncologie Médicale, EA 3738, CH Lyon-Sud, 69495 Pierre-Bénite Cedex, France [Gilles.Freyer@ chu-lyon.fr]

Lena E. Friberg, School of Pharmacy, University of Queensland, Brisbane 4072, Australia [l.friberg@pharmacy.uq.edu.au] *and* Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden

Stuart Friedrich, Global PK/PD and Trial Simulations, Eli Lilly Canada Inc., 3650 Danforth Ave., Toronto, ON, MIN 2E8 Canada [friedrich_stuart@Lilly.com]

Eliane Fuseau, EMF Consulting, Aix en Provence Cedex 4, France [eliane@emf-consulting.com]

Varun Garg, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 [varun_garg@vrtx.com]

Marc R. Gastonguay, Metrum Research Group LLC, 2 Tunxis Road, Suite 112, Tariffville, CT 06081 [marcg@metrumrg.com]

Panos G. Georgopoulos, Computational Chemodynamics Laboratory, Environmental and Occupational Health Sciences Institute, 70 Frelinghuysen Road, Piscataway, NJ 08854 [panosg@ccl.rutgers.edu]

Christopher J. Godfrey, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 *and* Anoixis Corp., 214 N. Main St., Natick, MA 01760 [cj-godfrey@anoixiscorp.com]

Thaddeus H. Grasela, Cognigen Corporation, 395 S Youngs Rd, Williamsville, NY 14221 [ted.grasela@cognigencorp.com]

David Hermann, deCODE Genetics, 1032 Karl Greimel Drive, Brighton, MI 48116 [david.hermann@decode.com]

Chuanpu Hu, Biostatistics, Sanofi-Aventis, 9 Great Valley Parkway, Malvern, PA 19355-1304 [Chuanpu.Hu@sanofi-aventis.com]

Matthew Hutmacher, Pfizer, PGR&D, 2800 Plymouth Road, Ann Arbor, MI 48105 [matt.hutmach er@pfizer.com]

Sastry S. Isukapalli, Computational Chemodynamics Laboratory, Environmental and Occupational Health Sciences Institute, 70 Frelinghuysen Road, Piscataway, NJ 08854 [ssi@fidelio.rutgers.edu]

Jin Y. Jin, Department of Pharmaceutical Sciences, School of Pharmacy, 519 Hochstetter Hall, State University of New York at Buffalo, Buffalo, NY 14260

Siv Jönsson, Clinical Pharmacology, AstraZeneca R&D Södertälje, SE-151 85 Södertälje, Sweden [siv.Jonsson@astrazeneca.com]

E. Niclas Jonsson, Hoffmann-La Roche Ltd., PDMP Modelling and Simulation, Grenzacherstr 124, Bldg. 15/1.052, CH-4070 Basel, Switzerland [niclas.jonsson@roche.com]

Karin Jorga, Hoffmann-La Roche Ltd., PDMP Clinical Pharmacology, Grenzacherstrasse 124, Bldg. 15/1.081A, CH-4070 Basel, Switzerland [karin.jorga@roche.com]

William J. Jusko, Department of Pharmaceutical Sciences, School of Pharmacy, 519 Hochstetter Hall, State University of New York at Buffalo, Buffalo, NY 14260 [wjjusko@buffalo.edu]

Mats O. Karlsson, Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden [mats.karlsson@farmbio.uu.se]

Ariya Khunvichai, Clinical Pharmacology, Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139 [ariya_khunvichai@vrtx.com]

Yong Ho Kim, Clinical Pharmacokinetics, Five Moore Drive, Sanders Bldg. 17.2245 PO Box 13398, Research Triangle Park, NC 27709 [joseph.y.kim@gsk.com] *and* Clinical Pharmacokinetics, GlaxoSmithKline, Raleigh, NC [yonghojoseph@hotmail.com]

Kenneth Kowalski, Pfizer, PGR&D, 2800 Plymouth Road, Ann Arbor, MI 48105 [ken.kowalski@pfizer.com]

James R. Lane, Department of Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 200 West Arbor Drive, San Diego, CA 92103-8765 [jrlane@ucsd.edu]

Donald E. Mager Department of Pharmaceutical Sciences, School of Pharmacy, 519 Hochstetter Hall, State University of New York at Buffalo, Buffalo, NY 14260 [dmager@acsu.buffalo.edu]

Mathilde Marchland, EMF Consulting, 425 rue Rene Descartes, BP 02, 13545 Aix-en-Provence Cedex 4, France [mathilde@emf-consulting.com]

Patrick J. Marroum, Office of Clinical Pharmacology, CDER, FDA, 10903 New Hampshire Avenue, Building 21, Silver Spring, MD 20993 [patrick.marroum@fda.hhs.gov]

Raymond Miller, Pfizer, PGR&D, 2800 Plymouth Road, Ann Arbor, MI 48105 [raymond.miller@pfizer.com]

Peter A. Milligan, Pfizer, Ramsgate Road, Sandwich, Kent, CT13 9NJ, UK [peter.a.milligan@pfizer.com]

Diane R. Mould, Projections Research, Inc., 535 Springview Lane, Phoenixville, PA 19460 [drmould@attglobal.net]

Partha Nandy, Johnson & Johnson Pharmaceutical Research and Development, 1125 Trenton-Hourborton Road, Titusville, NJ 08560 [pnandy@prdus.jnj.com]

Leonard C. Onyiah, Engineering and Computer Center, Department of Statistics and Computer Networking, St. Cloud State University, 720 4th Avenue South, St. Cloud, MN 56301 [lconyiah@stcloudstate.edu]

Olivier Pétricoul, EMF Consulting, 425 rue Rene Descartes, BP 02, 13545 Aix-en-Provence Cedex 4, France [olivier@emf-consulting.com]

José Pinheiro, Biostatistics, Novartis Pharmaceuticals Corporation, One Health Plaza, 419/2115, East Hanover, NJ 07936 [jose.pinheiro@novartis.com]

Murali Ramanathan, Pharmaceutical Sciences and Neurology, 543 Cooke Hall, State University of New York, Buffalo, NY 14260

Amit Roy, Strategic Modeling and Simulation, Bristol-Myers Squibb, Route 206 and Provinceline Road, Princeton, NJ 08540 [amit.roy@bms.com]

Matthew M. Riggs, Metrum Research Group LLC, 2 Tunxis Road, Suite 112, Tariffville, CT 06081 [mattr@metrumrg.com]

Mark E. Sale, Next Level Solutions LLC, 1013 Dickinson Circle, Raleigh, NC 27614 [mark@nextlevelsolns.com, msale2@yahoo.com]

He Sun, SunTech Research Institute, 1 Research Court, Suite 450-54, Rockville, MD 20850 [henrysunusa@yahoo.com, henrysunusa@gmail.com]

Brigitte Tranchand, Ciblage Thérapeutique en Oncologie, Faculté de Médecine, EA3738, Lyon-Sud, BP12, 69921 Oullins Cedex, France [Brigitte.Tranchand@adm.univ-lyon1.fr]

James A. Uchizono, Department of Pharmaceutics and Medicinal Chemistry, Thomas J. Long School of Pharmacy, University of the Pacific, Stockton, CA 95211 [juchizono@pacific.edu]

Paul J. Williams, Thomas J. Long School of Pharmacy and Health Sciences, University of the Pacific, Stockton, CA 95211 *and* Anoixis Corp., 1918 Verdi Ct., Stockton CA 95207 [pwilliams@pacific.edu, pkman@inreach.com]

Gary L. Wolk, 1215 South Kihei Rd., Kihei, HI 96753 [gwolk@hawaii.rr.com]

Jiuhong Zha, Biopharmacentical Sciences, Astellas Pharma, US Inc., Chicago [jiuhong.zha@us.astellas.com]

The subspecialty of population pharmacokinetics was introduced into clinical pharmacology / pharmacy in the late 1970s as a method for analyzing observational data collected during patient drug therapy in order to estimate patient-based pharmacokinetic parameters. It later became the basis for dosage individualization and rational pharmacotherapy. The population pharmacokinetics method (i.e., the population approach) was later extended to the characterization of the relationship between pharmacokinetics and pharmacodynamics, and into the discipline of pharmacometrics. Pharmacometrics is the science of interpreting and describing pharmacology in a quantitative fashion. Vast amounts of data are generated during clinical trials and patient care, and it is the responsibility of the pharmacometrician to extract the knowledge embedded in the data for rational drug development and pharmacotherapy. He/she is also responsible for providing that knowledge for decision making in patient care and the drug development process.

With the publication of the *Guidance for Industry: Population Pharmacokinetics* by the Food and Drug Administration (the advent of population pharmacokinetics/pharmacodynamics-based clinical trial simulation) and recently the *FDA Critical Path Initiative—The Critical Path to New Medical Products*, the assimilation of pharmacometrics as an applied science in drug development and evaluation is increasing. Although a great deal has been written in the journal literature on population pharmacokinetics, population pharmacokinetics/pharmacodynamics, and pharmacometrics in general, there is no major reference textbook that pulls all these facets of knowledge together in one volume for pharmacometricians in academia, regulatory agencies, or industry and graduate students/postdoctoral fellows who work/research in this subject area. It is for this purpose that this book is written.

Although no book can be complete in itself, what we have endeavored to assemble are contributors and an array of topics that we believe provide the reader with the knowledge base necessary to perform pharmacometric analysis, to interpret the results of the analysis, and to be able to communicate the same effectively to impact mission-critical decision making. The book is divided into seven sections—general principles, population pharmacokinetic basis of pharmacometrics, pharmacokinetics/pharmacodynamics relationship, clinical trial designs, pharmacometric knowledge creation, pharmacometric service and communication, and specific application examples. In the introductory chapter, the history of the development of pharmacometrics is traced and its application to drug development, evaluation, and pharmacotherapy is delineated. This is followed by Part I on general principles that addresses topics such as the general principles of programming, which is a must for every pharmacometrician, pharmacometric analysis software validation—a subject that has become prominent in last few years, linear and nonlinear mixed effects

modeling to provide the reader with the background knowledge on these topics and thus setting the pace for the remainder of the book, estimation of the dynamics of compliance, which is important for having a complete picture of a study outcome, graphical display of population data—a sine qua non for informative pharmacometric data analysis, the epistemology of pharmacometrics, which provides a pathway for performing a pharmacometric analysis, and data imputation. Data analysis without the proper handling of missing data may result in biased parameter estimates. The chapter on data imputation covers the various aspects of "missingness" and includes an example of how to handle left censored data—a challenge with most pharmacokinetic data sets.

In Part II of the book various aspects of population pharmacokinetics, pharmacometric knowledge discovery, and resampling techniques used in pharmacometric data analysis are covered. Thus, various aspects of the informative design and analysis of population pharmacokinetic studies are addressed together with population pharmacokinetics estimation methods. The chapter on pharmacometric knowledge discovery deals with the integrated approach for discovering knowledge from clinical trial data sets and communicating the same for optimal pharmacotherapy and knowledge/model-based drug development.

Part III, which is on the pharmacokinetics–pharmacodynamics relationship, deals with biomarkers and surrogates in drug development, gene expression analysis, integration of pharmacogenomics into pharmacokinetics/pharmacodynamics, empirical and mechanistic PK/PD models, outcome models, and disease progression models that are needed for understanding disease progression as the basis for building models that can be used in clinical trial simulation.

Part IV builds on the knowledge gained from the previous sections to provide the basis for designing clinical trials. The section opens with a chapter on the design of first-time-in-human (FTIH) studies for nononcology indications. The literature is filled with a discussion of the design of FTIH oncology studies, but very little has been written on the design of FTIH studies for nononcology indications. A comprehensive overview of different FTIH study designs is provided with an evaluation of the designs that provide the reader with the knowledge needed for choosing an appropriate study design. A comprehensive coverage of the design of Phase 1 and phase 2a oncology studies is provided in another chapter; the section closes with a chapter on the design of dose – response studies.

Part V addresses pharmacometric knowledge creation, which entails the application of pharmacometric methodologies to the characterization of an unexplored region of the response surface. It is the process of building upon current understanding of data that is already acquired by generating more data (information) that can be translated into knowledge. Thus, the section opens with a chapter on knowledge creation, followed by the theory of clinical trial simulation and the basics of clinical trial simulation, and ends with a chapter on the simulation of efficacy trials.

Parts VI and VII discuss what a pharmacometric service is all about, how to communicate the results of a pharmacometric analysis, and specific examples ranging from applications in a regulatory setting, characterization of QT interval prolongation, pharmacometrics in biologics development, pharmacometrics in pediatric pharmacotherapy, application of pharmacometric principles to the analysis of preclinical data, physiologically based pharmacokinetic modeling, characterizing metabolic and nonlinear pharmacokinetics, in vitro in vivo correlation, and the

application of pharmacometric knowledge discovery and creation to the characterization of drug safety.

What makes this book unique is not just the presentation of theory in an easy to comprehend fashion, but the fact that for a majority of the chapters there are application examples with codes in NONMEM, S-Plus, WinNonlin, or Matlab. The majority of the codes are for NONMEM and S-Plus. Thus, the reader is able to reproduce the examples in his/her spare time and gain an understanding of both the theory and principles of pharmacometrics covered in a particular chapter. A reader friendly approach was taken in the writing of this book. Although there are many contributors to the book, we have tried as much as possible to unify the style of presentation to aid the reader's understanding of the subject matter covered in each chapter. Emphasis has been placed on drug development because of the need to apply pharmacometrics in drug development to increase productivity. Examples have been provided for the application of pharmacometrics in pharmacotherapy and drug evaluation to show how pharmacometric principles have been applied in these areas with great benefit.

In the writing of this text, the reader's knowledge of pharmacokinetics, pharmacodynamics, and statistics is assumed. If not, the reader is referred to *Applied Pharmacokinetics* by Shargel and Yu, *Pharmacokinetics* by Gibaldi and Perrier, *Pharmacokinetics and Pharmacodynamics* by Gabrielson and Weiner, and statistics from standard textbooks.

Finally, this book is written for the graduate students or postdoctoral fellows who want to specialize in pharmacometrics; and for pharmaceutical scientists, clinical pharmacologists/pharmacists, and statisticians in academia, regulatory bodies, and the pharmaceutical industry who are in pharmacometrics or are interested in developing their skill set in the subject.

ENE I. ETTE
PAUL J. WILLIAMS

ACKNOWLEDGMENTS

Pharmacometrics: Impacting Drug Development and Pharmacotherapy

PAUL J. WILLIAMS and ENE I. ETTE

1.1 INTRODUCTION

Drug development continues to be expensive, time consuming, and inefficient, while pharmacotherapy is often practiced at suboptimal levels of performance (1–3). This trend has not waned despite the fact that massive amounts of drug data are obtained each year. Within these massive amounts of data, knowledge that would improve drug development and pharmacotherapy lays hidden and undiscovered. The application of pharmacometric (PM) principles and models to drug development and pharmacotherapy will significantly improve both (4, 5). Furthermore, with drug utilization review, generic competition, managed care organization bidding, and therapeutic substitution, there is increasing pressure for the drug development industry to deliver high-value therapeutic agents.

The Food and Drug Administration (FDA) has expressed its concern about the rising cost and stagnation of drug development in the white paper *Challenge and Opportunity on the Critical Path to New Products* published in March of 2004 (3). In this document the FDA states: "Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated in faster time frames, with more certainty, and at lower costs. . . . A new product development toolkit—containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques—is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product. We need superior product development science to address these challenges." In the critical path document, the FDA states that the three main areas of the path that need to be addressed are tools for assessing safety, tools for demonstrating medical utility, and lastly tools for characterization and manufacturing. Pharmacometrics can be applied to and can impact the first two areas, thus positively impacting the critical path.

For impacting safety, the FDA has noted opportunities to better define the importance of the QT interval, for improved extrapolation of in vitro and animal data to humans, and for use of extant clinical data to help construct models to screen candidates early in drug development (e.g., liver toxicity). Pharmacometrics can have a role in developing better links for all of these models.

For demonstrating medical utility, the FDA has highlighted the importance of model-based drug development in which pharmacostatistical models of drug efficacy and safety are developed from preclinical and available clinical data. The FDA goes on to say that "Systematic application of this concept to drug development has the potential to significantly improve it. FDA scientists use and are collaborating with others in the refinement of quantitative clinical trial modeling using simulation software to improve trial design and to predict outcomes." The pivotal role of pharmacometrics on the critical path is obvious.

Drug development could be improved by planning to develop and apply PM models along with novel pathways to approval, improved project management, and improved program development. Recent advances in computational speed, novel models, stochastic simulation methods, real-time data collection, and novel biomarkers all portend improvements in drug development.

Dosing strategy and patient selection continue to be the most easily manipulated parts of a patient's therapy. Optimal dosing often depends on patient size, sex, and renal function or liver function. All too often, the impact of these covariates on a PM parameter is unstudied and therefore cannot be incorporated into any therapeutic strategy. PM model development and application will improve both drug development and support rational pharmacotherapy.

1.2   PHARMACOMETRICS DEFINED

Pharmacometrics is the science of developing and applying mathematical and statistical methods to characterize, understand, and predict a drug's pharmacokinetic, pharmacodynamic, and biomarker–outcomes behavior (6). Pharmacometrics lives at the intersection of pharmacokinetic (PK) models, pharmacodynamic (PD) models, pharmacodynamic-biomarker–outcomes link models, data visualization (often by employing informative modern graphical methods), statistics, stochastic simulation, and computer programming. Through pharmacometrics one can quantify the uncertainty of information about model behavior and rationalize knowledge-driven decision making in the drug development process. Pharmacometrics is dependent on knowledge discovery, the application of informative graphics, understanding of biomarkers/surrogate endpoints, and knowledge creation (7–10). When applied to drug development, pharmacometrics often involves the development or estimation of pharmacokinetic, pharmacodynamic, pharmcodynamic–outcomes linking, and disease progression models. These models can be linked and applied to competing study designs to aid in understanding the impact of varying dosing strategies, patient selection criteria, differing statistical methods, and different study endpoints. In the realm of pharmacotherapy, pharmacometrics can be employed to customize patient drug therapy through therapeutic drug monitoring and improved population dosing strategies. To contextualize the role of pharmacometrics in drug development and pharmacotherapy, it is important to examine

the history of pharmacometrics. The growth of pharmacometrics informs much on its content and utility.

1.3   HISTORY OF PHARMACOMETRICS

1.3.1   Pharmacokinetics

Pharmacometrics begins with pharmacokinetics. As far back as 1847, Buchanan understood that the brain content of anesthetics determined the depth of narcosis and depended on the arterial concentration, which in turn was related to the strength of the inhaled mixture (11). Interestingly, Buchanan pointed out that rate of recovery was related to the distribution of ether in the body. Though there was pharmacokinetic (PK) work done earlier, the term pharmacokinetics was first introduced by F. H. Dost in 1953 in his text, *Der Blutspeigel-Kinetic der Knozentrationsablaufe in der Kreislauffussigkeit* (12). The first use in the English language occurred in 1961 when Nelson published his "Kinetics of Drug Absorption, Distribution, Metabolism, and Excretion" (13). The exact word pharmacokinetics was not used in this publication.

In their classic work, the German scientists Michaelis and Menton published their equation describing enzyme kinetics in 1913 (14). This equation is still used today to describe the kinetics of drugs such as phenytoin. Widmark and Tandberg (15) published the equations for the one-compartment model in 1924 and in that same year Haggard (16) published his work on the uptake, distribution, and elimination of diethyl ether. In 1934 Dominguez and Pomerene (17) introduced the concept of volume of distribution, which was defined as "the hypothetical volume of body fluid dissolving the substance at the same concentration as the plasma. In 1937 Teorrel (18) published a seminal paper that is now considered the foundation of modern pharmacokinetics. This paper was the first physiologically based PK model, which included a five-compartment model. Bioavailability was introduced as a term in 1945 by Oser and colleagues (19), while Lapp (20) in France was working on excretions kinetics.

Polyexponential curve fitting was introduced by Perl in 1960 (21). The use of analog computers for curve fitting and simulation was introduced in 1960 by two groups of researchers (22, 23).

The great growth period for pharmacokinetics was from 1961 to 1972, starting with the landmark works of Wagner and Nelson (24). In 1962 the first symposium with the title pharmacokinetics, "Pharmacokinetik und Arzniemitteldosireung," was held.

Clinical pharmacokinetics began to be recognized in the 1970s, especially in two papers by Gibaldi and Levy, "Pharmacokinetics in Clinical Practice," in the *Journal of the American Medical Association* in 1976 (25). Of further importance that same year was a paper by Koup et al. (26) on a system for the monitoring and dosing of theophylline based on pharmacokinetic principles.

Rational drug therapy is based on the assumption of a causal relationship between exposure and response. There pharmacokinetics has great utility when linked to pharmacodynamics and the examination of pharmacodynamics is of paramount importance.

1.3.2 Pharmacodynamics

In 1848 Dungilson (27) stated that pharmacodynamics was "a division of pharmacology which considers the effects and uses of medicines." This definition has been refined and restricted over the centuries to a more useful definition, where "pharmacokinetics is what the body does to the drug; pharmacodynamics is what the drug does to the body" (28, 29). More specifically, pharmacodynamics was best defined by Derendorf et al. (28) as "a broad term that is intended to include all of the pharmacological actions, pathophysiological effects and therapeutic responses both beneficial or adverse of active drug ingredient, therapeutic moiety, and/or its metabolite(s) on various systems of the body from subcellular effects to clinical outcomes." Pharmacodynamics most often involves mathematical models, which relate some concentration (serum, blood, urine) to a physiologic effect (blood pressure, liver function tests) and clinical outcome (survival, adverse effect). The pharmacodynamic (PD) models have been described as fixed, linear, log-linear, E_{max}, sigmoid E_{max}, and indirect PD response (29–31).

The indirect PD response model has been a particularly significant contribution to PD modeling (30, 31). It has great utility because it is more mechanistic than the other models, does not assume symmetry of the onset and offset, and incorporates the impact of time in addition to drug concentration, thus accounting for a delay in onset and offset of the effect. For these models the maximum response occurs later than the time of occurrence of the maximum plasma concentration because the drug causes incremental inhibition or stimulation as long as the concentration is "high enough." After the response reaches the maximum, the return to baseline is a function of the dynamic model parameters and drug elimination. Thus, there is a response that lasts beyond the presence of effective drug levels because of the time needed for the system to regain equilibrium. Whenever possible, these mechanistic models should be employed for PD modeling and several dose levels should be employed for accurate determination of the PD parameters, taking into consideration the resolution in exposure between doses.

The dependent variables in these PD models are either biomarkers, surrogate endpoints, or clinical endpoints. It is important to differentiate between these and to understand their relative importance and utility.

1.3.3 Biomarkers

The importance of biomarkers has been noted in recent years and is evidenced by the formation of The Biomarkers Definitions Working Group (BDWG) (32). According to the BDWG, a biomarker is a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic process or pharmacologic responses to a therapeutic intervention." Biomarkers cannot serve as penultimate clinical endpoints in confirming clinical trials; however, there is usually considered to be some link between a biomarker based on prior therapeutic experience, well understood physiology or pathophysiology, along with knowledge of the drug mechanism. Biomarkers often have the advantage of changing in drug therapy prior to the clinical endpoint that will ultimately be employed to determine drug effect, thus providing evidence early in clinical drug development of potential efficacy or safety.

A surrogate endpoint is "a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit, harm, lack of benefit, or lack of harm based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence" (32). Surrogate endpoints are a subset of biomarkers such as viral load or blood pressure. All surrogate endpoints are biomarkers. However, few biomarkers will ever become surrogate endpoints. Biomarkers are reclassified as surrogate endpoints when a preponderance of evidence indicates that changes in the biomarker correlate strongly with the desired clinical endpoint.

A clinical endpoint is "a characteristic or variable that reflects how a patient feels, functions or survives. It is a distinct measurement or analysis of disease characteristics observed in a study or a clinical trial that reflect the effect of a therapeutic intervention. Clinical endpoints are the most credible characteristics used in the assessment of the benefits and risks of a therapeutic intervention in randomized clinical trials." There can be problems with using clinical endpoints as the final measure of patient response because a large patient sample size may be needed to determine drug effect or the modification in the clinical endpoint for a drug may not be detectable for several years after the initiation of therapy.

There are several ways in which the discovery and utilization of biomarkers can provide insight into the drug development process and patient care. Biomarkers can identify patients at risk for a disease, predict patient response, predict the occurrence of toxicity, and predict exposure to the drug. Given these uses, biomarkers can also provide a basis for selecting lead compounds for development and can contribute knowledge about clinical pharmacology. Therefore, biomarkers have the potential to be one of the pivotal factors in drug development—from drug target discovery through preclinical development to clinical development to regulatory approval and labeling information, by way of pharmacokinetic/pharmacodynamic–outcomes modeling with clinical trial simulations.

1.3.4 PK/PD Link Modeling

PK/PD modeling provides the seamless integration of PK and PD models to arrive at an enlightened understanding of the dose–exposure–response relationship. PK/PD modeling can be done either sequentially or simultaneously (33, 34). Sequential models estimate the pharmacokinetics first and fix the PK parameters, generating concentrations corresponding to some PD measurement. Thus, the pharmacodynamics is conditioned on the PK data or on the estimates of the PK parameters. Simultaneous PK/PD modeling fits all the PK and PD data at once and the PK and PD parameters are considered to be jointly distributed. When simultaneous modeling is done, the flow of information is bidirectional. Both of these approaches appear to provide similar results (33, 35). However, it is important to note that PD measurements are usually less precise than PK measurements and using sequential PK and PD modeling may be the preferred approach in most instances.

PK and PD can be linked directly through a measured concentration that is directly linked to an effect site. The direct link model does not work well when there is a temporal relationship between a measured concentration and effect, as when hysteresis is present. When this is the case, an indirect link between the measured concentration and effect must be accounted for in the model. This has been done in

general by the construction of an effect compartment, where a hypothetical effect compartment is linked to a PK compartment. Here the effect compartment is very small and thus has negligible impact on mass balance with a concentration time course in the effect compartment. The effect is related to the concentration in the effect compartment, which has a different time course than the compartment where drug concentrations are actually measured. In addition to the effect compartment approach to account for temporal concentration–effect relationships, the indirect response concept has found great utility.

PK and PD have been linked by many models, sometimes mechanistic and at other times empirical. These models are especially useful in better understanding the dose strategy and response, especially when applied by stochastic simulation. The population approach can be applied to multiple types of data—for example, both intensely and sparsely sampled data and preclinical to Phase 4 clinical data—and therefore has found great utility when applied to PK/PD modeling.

1.3.5   Emergence of Pharmacometrics

The term pharmacometrics first appeared in the literature in 1982 in the *Journal of Pharmacokinetics and Biopharmaceutics* (36). At that time, the journal made a commitment to a regular column dealing with the emerging discipline of pharmacometrics, which was defined as "the design, modeling, and analysis of experiments involving complex dynamic systems in the field of pharmacokinetics and biopharmaceutics . . . concerning primarily data analysis problems with such models." They went on to say that problems with study design, determination of model identifiability, estimation, and hypothesis testing would be addressed along with identifying the importance of graphical methods. Since this time, the importance of pharmacometrics in optimizing pharmacotherapy and drug development has been recognized, and several graduate programs have been established that emphasize pharmacometrics (37). Pharmacometrics is therefore the science of developing and applying mathematical and statistical methods to (a) characterize, understand, and predict a drug's pharmacokinetic and pharmacodynamic behavior; (b) quantify uncertainty of information about that behavior; and (c) rationalize data-driven decision making in the drug development process and pharmacotherapy. In effect, pharmacometrics is the science of quantitative pharmacology.

1.3.6   Population Modeling

A major development in pharmacometrics was the application of population methods to the estimation of PM parameters (38). With the advent of population approaches, one could now obtain estimates of PM parameters from sparse data from large databases and also obtain improved estimates of the random effects (variances) in the parameters of interest. These models first found great applicability by taking massive amounts of data obtained during therapeutic drug monitoring (TDM) from which typical values and variability of PK parameters were obtained. The parameters once estimated were applied to TDM to estimate initial doses and, using Bayesian algorithms, to estimate a patient's individual PK parameters to optimize dosing strategies. Population methods have become widely accepted to the

extent that a *Guidance for Industry* has been issued by the United States Food and Drug Administration (FDA) on population pharmacokinetics. Population methods are applied to pharmacokinetics, pharmacodynamics, and models linking biomarkers to clinical outcomes (39).

1.3.7 Stochastic Simulation

Stochastic simulation was another step forward in the arena of pharmacometrics. Simulation had been widely used in the aerospace industry, engineering, and econometrics prior to its application in pharmacometrics. Simulation of clinical trials first appeared in the clinical pharmacology literature in 1971 (40) but has only recently gained momentum as a useful tool for examining the power, efficiency, robustness, and informativeness of complex clinical trial structure (41).

A major impetus promoting the use of clinical trial simulation was presented in a publication by Hale et al. (41), who demonstrated the utility of simulating a clinical trial on the construction of a pivotal study targeting regulatory approval. The FDA has shown interest in clinical trial simulation to the extent that it has said: "Simulation is a useful tool to provide convincing objective evidence of the merits of a proposed study design and analysis. Simulating a planned study offers a potentially useful tool for evaluating and understanding the consequences of different study designs" (39). While we often think of clinical trial simulation as a way for the drug sponsor to determine optimal study structure, it is also a way for the FDA to determine the acceptability of a proposed study protocol. Simulation serves as a tool not only to evaluate the value of a study structure but also to communicate the logical implications of a PM model, such as the logical implication of competing dosing strategies for labeling.

The use and role of a simulated Phase 3 safety and efficacy study is still under discussion as confirmatory evidence at the FDA; however, a simulation of this type can serve as supportive evidence for regulatory review (4, 5). It is likely that at some time in the future knowledge of a disease's pathophysiology plus knowledge of drug behavior and action will be applied to a group of virtual patients as the pivotal Phase 3 study for approval by a clinical trial simulation. Stochastic simulation should result in more powerful, efficient, robust, and informative clinical trials; therefore, more can be learned, and confirming efficacy will be more certain as stochastic simulation is applied to the drug development process.

1.3.8 Learn–Confirm–Learn Process

Drug development has traditionally been empirical and proceeded sequentially from preclinical through clinical Phases 1 to 3. Sheiner (42) first proposed a major paradigm shift in drug development away from an empirical approach to the learn–confirm approach based on Box's inductive versus deductive cycles (43). Williams et al. (6, 44) and Ette et al. (45) have since revised this process to the learn–confirm–learn approach because of their emphasis on the fact that learning continues throughout the entire drug development process. The learn–confirm–learn process contends that drug development ought to consist of alternate cycles of learning from experience and then confirming what has been learned but that one never proposes a protocol where learning ceases.

In the past, Phases 1 and 2a have been considered the learning phases of drug development because the primary objectives are to determine the tolerated doses and the doses producing the desired therapeutic effect. Phase 2 has targeted how to use the drug in the target patient population, determining the dose strategy and proof of concept. Phase 3 has focused on confirming efficacy and demonstrating a low incidence of adverse events, where if the ratio of benefit to risk is acceptable then the drug is approved. An encouraging outcome in these early cycles results in investment in the costly Phase 2b and 3 studies. However, even in the confirming stages of drug development, one ought to continue to be interested in learning even though confirming is the primary objective of a study; that is, all studies should incorporate an opportunity for learning in the protocol. Therefore, the process has been renamed "learn–confirm–learn".

Learning and confirming have quite different goals in the process of drug development. When a trial structure optimizes confirming, it most often imposes some restrictions on learning; for example, patient enrollment criteria are limited, thus limiting one's ability to learn about the agent in a variety of populations. For example, many protocols limit enrollment to patients with creatinine clearances above a certain number (e.g., 50 mL/min). If this is done, one cannot learn how to use such a drug in patients with compromised renal function. Empirical commercial drug development has in general focused on confirming because it provides the necessary knowledge for regulatory approval, addressing the primary issue of efficacy. The downside of the focus on confirming is that it has led to a lack of learning, which can result in a dysfunctional drug development process and less than optimal pharmacotherapy postapproval.

PM modeling focuses on learning, where the focus is on building a model that relates dosing strategy, exposure, patient type, prognostic variables, and more to outcomes. Here the three-dimensional response surface is built (42) (see Section 1.3.9.2). PM models are built to define the response surface to increase the signal-to-noise ratio, which will be discussed shortly. The entire drug development process is an exercise of the learn–confirm–learn paradigm.

1.3.9 Exposure–Response Relationship

The importance of elucidating the exposure–response relationship must be emphasized. When the term exposure is used, one is usually referring to dose or variables related to concentration such as area under the concentration–time curve (AUC), maximum concentration (C_{max}), minimum concentration (C_{min}), or average concentration (C_{ave}) in some biological specimen such as serum, urine, cerebral spinal fluid, or sputum. It is worth noting that dose is a very weak surrogate of exposure, especially where there is no proportionality between dose and AUC or C_{max}. Response is a measure of the effect of a drug either therapeutic or adverse, such as blood pressure, cardiac index, blood sugar, survival, liver function, or renal function.

1.3.9.1 Regulatory Perspective

The FDA document, *Guidance for Industry: Exposure–Response Relationships— Study Design, Data Analysis, and Regulatory Applications*, has commented extensively on the exposure–response relationship (46). It states: "Exposure–response information is at the heart of any determination of the safety and effectiveness of

drugs. . . . In most cases, however, it is important to develop information on the population exposure–response relationships for favorable and unfavorable effects and information on how, and whether, exposure can be adjusted for various subsets of the population." The FDA recognizes the value of exposure–response knowledge to support the drug development process and to support the determination of safety and efficacy. In this document it stated that "dose–response studies can, in some cases, be particularly convincing and can include elements of consistency that, depending on the size of the study and outcome, can allow reliance on a single clinical efficacy study as evidence of effectiveness." The exposure–response relationship was further refined in the defining of the response surface.

1.3.9.2 *Response Surface*

A significant development of the exposure–response concept was the proposing of the response surface. Sheiner (42) first proposed the pharmacological response surface as a philosophical framework for development of PM models. The response surface can be thought of as three dimensional: on one axis are the input variables (dose, concurrent therapies, etc.); on the second axis are the important ways that patients can differ from one another that affect the benefit to toxicity ratio; and the final axis represents the benefit to toxicity ratio. Sheiner stated: "the real surface is neither static, nor is all the information about the patient conveyed by his/her initial prognostic status, nor are exact predictions possible. A realistically useful response . . . must include the elements of variability, uncertainty and time . . ." Thus, the primary goal of the response model is to define the complex relationship between the input profile and dose magnitude when comparing beneficial and harmful pharmacological effects and how this relationship varies between patients. For rational drug use and drug development, the response surface must be mapped. PM models, once developed and validated, allow extrapolation beyond the immediate study subjects to allow application to other patients from whom the model was not derived. These predictive models permit the evaluation of outcomes of competing dosing strategies in patients who have not received the drug and therefore aid in constructing future pivotal studies. One important aspect of PM models employed in mapping the response surface is that they increase the signal-to-noise ratio in a data set because they translate some of the noise into signal. This is important because when we are converting information (data) into knowledge, the knowledge is proportional to the signal-to-noise ratio.

1.3.10 PM Knowledge Discovery

It is our experience that most drug development programs are data rich and knowledge poor. This occurs when data are collected but all of the knowledge hidden in the data set is not extracted. In reality, huge amounts of data are generated from modern clinical trials, observational studies, and clinical practice, but at the same time there is an acute widening gap between data collection, knowledge, and comprehension. PM knowledge discovery applies 13 comprehensive and interwoven steps to PM model development and communication and relies heavily on modern statistical techniques, modern informative graphical applications, and population modeling (8, 9) (see Chapter 14). The more that is known about a drug the better will be its application to direct patient care, and the more powerful and efficient

will be the development program. To this end, PM knowledge discovery is the best approach to extracting knowledge from data and has been defined and applied to PM model development.

1.3.11   PM Knowledge Creation

Most often, knowledge discovery provides the foundation for knowledge creation and is simply the initial step in the application of PM knowledge (10). The discovered knowledge can be used to synthesize new data or knowledge, or to supplement existing data. PM knowledge creation has something in common with knowledge discovery its intent to understand and better define the response surface. Data supplementation deals with the use of models on available data to generate supplemental data that would be used to characterize a targeted unexplored segment of the response surface (47).

1.3.12   Model Appropriateness

Model appropriateness brought a new epistemology to PM model estimation and development (48) (see Chapter 8). The pivotal event in establishing model appropriateness is stating the intended use of the model. The entire process requires the stating of the intended use of the model, classifying the model as either descriptive or predictive, evaluating the model, and validating the model if the model is to be used for predictive purposes. Descriptive models are not intended to be applied to any external population—that is, their sole purpose is to gain knowledge about the drug in the population studied. Predictive models are intended to be applied to subjects from whom the model was not derived or estimated. Predictive models require a higher degree of correspondence to the external universe than descriptive models and therefore require validation.

Under the epistemology of model appropriateness, the purpose for which the model is developed has a significant impact on the modeling process. In the current modeling climate, insufficient consideration is given to the purpose or intended use of the model and little attention is given to whether the model is descriptive or predictive. Model appropriateness is a paradigm that ought to be applied to the model development and estimation process and it provides the framework for appropriate use of PM models.

1.4   PIVOTAL ROLE OF PHARMACOMETRICS IN DRUG DEVELOPMENT

Drug development has become protracted and expensive over the last several decades, with the average length of clinical development being over 7–12 years, the number of studies averaging 66, and a cost of $0.802–1.7 billion per approved agent (1–4). The process has been empirical—driven by identifying all the items needed for registration of an agent, constructing a checkbox for each item, and executing the studies so that each box is checked, with a consequent fulfillment of each requirement. The numbers above indicate that this empirical, "it has always been done this way" approach does not work well and novel approaches need to be applied. The learn–confirm–learn paradigm should be applied to all drug

development programs, and modeling should follow the epistemology of model appropriateness.

To expedite drug development while maintaining patient safety, new technologies and approaches to discovery, improved project and development approaches, portfolio review, application of sound science, novel study structures, and pharmacometrically guided development programs will need to emerge (49). The use of pharmacometrics to define the dose exposure–response relationship has been successful in improving drug development and pharmacotherapy. Of pivotal importance here is the learn–confirm–learn paradigm, which has been previously mentioned as one of the significant proposals in the evolution of pharmacometrics.

While pharmacometrics can be an important tool to expedite drug development, it will also play a key role in determining the optimal dose at the time of approval (new drug application approval). Going to market with the optimal dose is not as straightforward as one may expect. A recent retrospective study noted that of 499 approved drugs between 1980 and 1999, one in five had a dosage change postapproval and 80% of these changes were a decrease in dose (50). This study concluded that current drug development frequently does not capture completely the dose information needed for safe pharmacotherapy. To address this, Cross et al. (50) suggested that improved PK and PD information be gathered early in Phase 2 of drug development. Finally, if drug doses are higher than need be during development and adverse events are related to dose, this may result in an increased frequency of adverse events resulting in an increased study dropout rate and therefore a decrease in study power.

Finding the optimal dose is one of the primary goals of clinical development, because changing a dose based on patient characteristics can easily be done. Simplified dosing strategies are often sought by the drug sponsor because it results in ease of use by the practitioner and the patient. Often a sponsor wants a "one dose fits all" approach, which may not result in optimized dosing. Often several levels of dose stratification result in surprisingly improved dosing strategies (e.g., elderly versus young).

Novel study structures, such as the enrichment trial, fusion, and adaptive design studies, will result in more efficient drug development. Enrichment studies attempt to choose subjects who are likely to respond. Study groups can be "enriched" by enrolling only subjects with response markers in a specific range or by enrolling only subject types demonstrating a good response during a short pretest phase. In enrichment trials the exposure relationship can be studied efficiently, but it is difficult to know how to extrapolate the quantitative relationship (exposure–response) from an enrichment study to the general population.

The advantage of the adaptive design study is that it emphasizes study of the drug in the region of useful doses, thus minimizing the number of subjects in regions where the drug is not effective. For adaptive designs, an exposure–response model is used and continuously updated as each subject's response is observed. The updated model is used to generate the probability of allocation of each new subject to a treatment arm, favoring the allocation to those arms with the better accumulated outcomes to date, with new subjects randomly allocated to arms on the basis of these frequencies. A treatment arm is dropped from the remainder of the study when its allocation probability drops below a specified threshold. The efficiency of this study design is that as few subjects as necessary are studied to determine that

one dose level is less useful than another. This approach can decrease study duration and numbers of subject in a clinical study. Adaptive design works best when patient accrual rates are slow.

1.4.1 Preclinical Development

Drug discovery has focused on identifying the most potent lead compound for a specified target. However, many drugs have failed due to poor pharmacokinetic or biopharmaceutical properties such as a short half-life or poor bioavailability. In today's economic environment such failures can no longer be afforded. It has become recognized that the "best drug" is one that balances potency, good pharmacokinetic–biopharmaceutical properties, good pharmacodynamics, safety, and low cost of manufacturing. It is important to deal with these issues prior to testing in humans.

Optimized preclinical development can be a tremendous aid to the design of early clinical studies. This optimization will include a thorough study of preclinical safety by combining traditional toxicology studies with novel methods in toxicoproteomics, toxicogenomics, and metabolomics. These new "-omics" will lead to novel biomarkers to predict toxicology and efficacy.

Preclinical development should play an important role in defining the exposure–response (both efficacy and toxicity) relationships, which is a primary role for preclinical pharmacometrics. It is essential to determine the absorption, distribution, metabolism, and elimination during toxicokinetic studies in order to understand the comparison of these across species. It has been demonstrated that by combining preclinical exposure–response data (the steepness of the curve is important here), preclinical pharmacokinetics, and novel approaches to scale up to humans (10, 51) (see also Chapters 29 and 30), Phase 1 can be expedited. This can be done by choosing higher first time in human doses or more rapid escalation (if the dose–response curve is rather flat), resulting in fewer dosing cycles and thus less time, energy, and finances expended on Phase 1, without sacrificing safety.

The development of physiologically and pathophysiologically based PM models (PBPM models) during preclinical development deserves attention. These models have the potential to provide accurate and nearly complete characterization of the PK and concentration–effect relationship and quantification of the potency of a drug (52–56). PBPM testing is best executed when the chemistry, biochemistry, metabolism, and exposure response of the drug are well known in addition to the relative physiology of the animals used in preclinical trials versus the parallel human physiology. To utilize PBPM modeling one must define the physiology, pathophysiology, biochemistry, and exposure–response relationships. To execute this type of modeling, some of the physiological variables that often need to be defined include blood flow to various organs such as liver, kidney, and effect organs. The biochemical–pharmacological parameters of a model that often need to be defined are K_m and V_{max} for the various enzymes that catalyze the metabolism of the drug and/or metabolites; tissue to blood concentration ratios; the distribution of the drug and/or metabolites of interest, for example, protein binding; and the clearance for various organs, for example, liver versus kidney. Exposure–response variables that are associated with a positive response or an adverse event need to be identified such as area under the concentration–time curve (AUC) or maximum concentra-

tion (C_{max}) or nadir concentration (C_{min}). The exposure response may be related to the parent compound or to a metabolite and may be a concentration-based variable in plasma or within a specific organ or tumor. Many of these parameters can be estimated in vitro, such as enzyme kinetic parameters and protein binding, and physiologic parameters can be obtained from the literature, such as blood flow rates and organ volumes (56).

PBPM modeling enabled the evaluation of the pharmacometrics of capecitabine for determination of the optimal dosing strategy in humans (56). Capecitabine is a prodrug that is converted in three steps to 5-fluorouracil (5-FU). A multicompartmental model was developed to describe the pharmacometrics of capecitabine, two metabolites, and 5-FU. The PBPM model is shown in Figure 1.1. The model included five compartments, all in some way related to either efficacy or adverse event. The parameters included in the model were K_m and V_{max} for each of the enzymes that catalyze capecitabine to 5-FU; tissue to blood ratio of capecitabine and the metabolites in gastrointestinal (GI), liver, and tumor tissue; protein binding; blood flow rate to liver, GI, and tumor tissue; and urinary clearance of unbound capecitabine and its metabolites. Enzyme activities (liver, breast, and colorectal tumors) and protein binding parameters were derived from in vitro experiments. Physiologic parameters were obtained from the literature.

From the model, the 5-FU AUC values in breast and colorectal tumors were simulated at doses from 829 to 1255 mg/m^2. The 5-FU AUC in tumor increased in a nonlinear manner relative to the increases in capecitabine dose. The model indicated that, for capecitabine, the 5-FU exposure in the tumors was much greater than in blood, resulting in a relatively low systemic exposure. The simulated blood
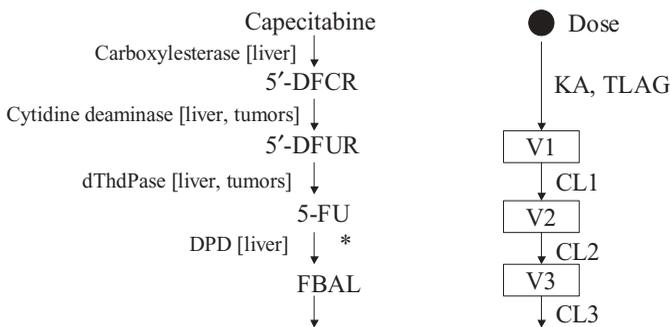
FIGURE 1.1 Metabolic pathway of capecitabine and its representation by a PK model. Abbreviations: Tissues with high enzyme activites are shown in square brackets; 5′-DFCR = 5′deoxy-5-flurocytidine; 5′-DFUR = 5′deoxy-5-flurouridine; dThdPase = thymidine phosphorylase; DPD = dihydropyrimidine dehydrogenase; FBAL = α-fluoro-β-alanine; FUH$_2$ = dihydro-5-fluorouracil; FUPA = 5-fluoro-ureido-propionic acid. Dose = capecitabine dose (mg); KA = first-order absorption rate constant (L/h); TLAG = lagtime (h); CL1 = apparent 5′-DFUR clearance (L/h); V1 = apparent 5′-DFUR volume (L); CL2 = apparent 5-FU clearance (L/h); V2 = apparent 5-FU volume (V); CL3 = apparent FBAL clearance (L/h); V3 = apparent FBAL volume (L). (From Blesch et al. (56); used with permission.)

AUC values were consistent with clinical observations, indicating that the model was able to describe known clinical data.

Once the model was developed, a murine xenograft was done and the PK, blood, and tissue binding of capecitabine and its metabolites were measured in vivo and integrated into the PBPM model. Large interspecies differences in tissue distribution and metabolic activity were observed. The predicted blood and tissue concentration profiles of 5-FU in the xenograft were compared to those in humans after simulated oral administration of several levels of capecitabine doses. The 5-FU *AUC*s in blood and xenograft tumor tissues were lower than those in humans for all capecitabine doses administered. At their effective oral doses of capecitabine (0.0944 mmol/kg, the clinical effective dose for humans; 0.44 mmol/kg, the effective dose for human cancer xenograft) similar 5-FU *AUC* values were observed in humans and human cancer xenograft models. The results of this study strongly supported the fact that a clinically effective dose can be extrapolated from xenograft models to a corresponding effect dose in humans when thoughtful approaches to the development and application of PBPM modeling is executed. Preclinical PM modeling should be done on a real-time basis so that modeling has been completed prior to planning and protocol development for Phase 1.

Biomarkers need to be identified and investigated in preclinical studies, especially those that may predict future safety problems. Sometimes the lowering of blood pressure or the prolongation of the corrected QT interval may give one a "heads up" to potential toxicities or dose-related toxicities that may occur during clinical development. When a thorough job is done during preclinical development, then transition to clinical development can be done efficiently and with confidence.

1.4.2 Clinical Development

Clinical development continues with the application of the learn–confirm–learn paradigm applied to drug development. Scale up to the first-time-in-human (FTIH) study is best done by the application of sound PM methods as described by several authors (10, 51, 56).

1.4.2.1 *Phase 1 Studies*
Phase 1 studies are executed to identify well tolerated doses and, in some cases, the maximum tolerated dose, to study the single and multiple dose pharmacokinetics, and to gain an initial knowledge of the exposure–response relationship. In addition to the above, one sometimes does Phase 1 studies to determine food effect and gender on pharmacokinetics, drug–drug interactions, and pharmacokinetics in special populations such as those with impaired renal or hepatic function or pediatric or geriatric patients. Here one has learned about the dose–exposure–response relationship from preclinical studies, has been guided by that preclinical knowledge, and is confirming or revising what was learned. Both traditional two-stage and population PK methods have been applied to Phase 1 model development with good results. The population approach can provide valuable information that is otherwise not available by the standard two-stage approach. Phase 1 studies are most often conducted in healthy volunteers unless the anticipated toxicity of the drug is severe or the drug is being applied to a life-threatening condition for which no other treatment is available.

In Phase 1, the approach to the FTIH study is critical in determining how much time is expended in this part of development. The central issue here is: "What should the first dose be and how rapidly does escalation occur?" If the very first dose it too high, then an adverse event will occur; if it is too low, then unnecessary time will be expended on low-dose testing. The application of preclinical findings becomes important. A promising approach has been the combining of allometry and mixed effect modeling with stochastic simulation to extrapolate preclinical models and knowledge to humans (10, 51). Applying sound PM methods has been and will be of great value in bringing efficiency to Phase 1 studies and for discovering knowledge that was previously hidden in most Phase 1 data sets. In situations where the maximum tolerated dose (MTD) is sought and defined in healthy volunteers, it should be redefined in patients at some later stage of development if possible (57, 58).

In addition to the FTIH studies, the effects of food, drug–drug interactions, and special populations need to be studied. Coadminstration of drugs has been demonstrated to both increase and decrease bioavailability of some agents with the subsequent lack of efficacy or appearance of toxicity. Further details on the design and conduct of food effect studies can be found in Chapter 29. Drug–drug interaction studies have become increasingly important as the number of agents prescribed to patients continues to increase. In one instance, a prominent drug was withdrawn from the market after adverse events were reported, which were due to interactions with other agents. It is important to obtain information for some subpopulations, such as pediatric patients, those with renal impairment, and the elderly, so that group-specific dosing guidelines can be developed. These special studies can be executed with either traditional PK studies or more efficiently by applying population techniques (39) (see Chapters 12 and 39). The need to study subpopulations strongly supports implementing the learn–confirm–learn paradigm. These issues are addressed in Chapter 14.

As the development process nears the end of Phase 1, it becomes crucial to extract all knowledge from existing data. PM models should be developed, linking drug exposure to pharmacodynamics (response). These models are applied, often by stochastic simulation, to optimize the structure and designs of Phase 2 studies. Real-time data collection is helpful here so that PM models may be estimated prior to data set closure and then applied to evaluation of competing Phase 2a study designs (39, 48, 59, 60). In this way, efficient and powerful Phase 2 programs can be constructed.

1.4.2.2 Phase 2 Studies

Phase 2 studies should focus on both learning and confirming. Historically, Phase 2a has had as its primary goal to demonstrate "proof of concept" that the drug is capable of being effective. It has been a common practice to administer the maximum tolerated dose (MTD) in Phase 2a and this dose may be on the flat part of the efficacy curve. If this is the case, lower doses may have been equally effective and less toxic. This dose is then carried forward into Phase 2b and eventually Phase 3. In Phase 3 the drug will likely be demonstrated to be effective and without significant adverse effects. The result will be NDA approval at the MTD. Therefore, doses may be lowered because "a lower dose is quite adequate for treatment and less expensive" in the opinion of the prescriber or "a lower safer dose may be

needed." The former may be enacted by practitioners without a change in labeling and the latter would come at the directive of the FDA. The former can be quite costly in terms of gross revenues for the manufacturer because an increase in cost per unit after marketing is in general not a viable alternative.

Phase 2a should have learning as its primary focus to define the optimal dose, thus improving the drug development process; while Phase 2b studies should focus on confirming. Phase 2a is the time during development to learn about efficacy; to confirm or modify what was learned in Phase 1 about safety, efficacy, and drug effect on biomarkers; and to refine the dose–PK/PD-biomarkers–surrogate–outcomes relationships.

The knowledge discovered in Phase 2a provides information for the later larger trials that will be designed to prove efficacy. The sample sizes are small in Phase 2 and the patients are often the "healthiest" to minimize disease-related variability. With this in mind, the Phase 2a study should be designed to give a first glimpse to the following issues (48): (a) Does the drug work? (b) How does the drug work? (c) What is the dose–response relationship? (d) Is there a difference in any of the pharmacology in subgroups? A very valuable practice here is to power these studies by setting α at a more liberal level of 0.10–0.20 when evaluating efficacy. Addressing these issues will require paying attention to important design points such as number and level of doses studied, timing of endpoint observations, number of subjects at each dosing level, and duration of the study. Furthermore, a well designed Phase 2a trial with 150–200 subjects will usually provide more information and is less costly than several smaller studies, even when these are later combined (48). A well designed study here will usually depend on stochastic simulation of competing study designs. In the end, many of the analyses will be population dose–pharmacokinetics/pharmacodynamics–response models.

In Phase 2 the proof of concept study provides scientifically sound evidence supporting the postulated effect of the new drug, where the effect may be the relevant pharmacological action or a change in disease biomarkers, established surrogate endpoints, or clinical outcomes that may be beneficial and/or toxic in nature. The proof of concept is often used for go/no-go decisions and is therefore one of the most critical steps in the drug development process.

Biomarkers play an important role in Phase 2 studies. These are covered in Chapter 20 in detail. Biomarkers are most important in early efficacy and toxicity studies when clinical endpoints take too long to become observable. After approval, biomarkers may prove useful in monitoring the course of pharmacotherapy in individual patients.

Prior to advancing to Phase 2b, all the knowledge hidden in the Phase 1 and Phase 2a data ought to be discovered. Then clinical trial simulation (knowledge creation) should be applied to construct Phase 2b.

In Phase 2b the knowledge discovered in all previous phases is confirmed, and learning more about the drug in a larger patient population continues. In this phase of development, strong supportive evidence is generated so that if an accelerated approval is sought the knowledge and data generated could be enough to obviate the need for two Phase 3 confirming studies. Attention should be given to informatively designing Phase 2b studies to meet the confirming study objectives and allow learning that will enhance a further characterization of the response surface. Pharmacokinetics enables the refinement and further development of PK/PD models

for dosage optimization (see Chapter 29). In Phase 2b sparse sampling is adequate; this data may be concatenated with previously collected data. The concatenation of these data with previously collected data and the estimation of individual PK or PD parameters via post hoc Bayesian algorithms may be useful for explaining individual treatment failures, toxicities, or positive responses to a drug. The PM models estimated from all previous data and available at the end of Phase 2b are important for constructing the pivotal Phase 3 program through knowledge creation.

1.4.2.3   Phase 3

Phase 3 is the pivotal phase for registration of a drug, where usually two large randomized, controlled trials for establishing efficacy and safety are required. The PM models from all previous studies are crucial for the determination of the dose(s), patient population selection, study duration, number of patients, and so on for Phase 3. In some cases a single pivotal study may be acceptable to the regulatory agency provided there is good supportive science (which may be good PM models) and confirmatory evidence supporting efficacy and safety (6, 7). In Phase 3 it is still advisable to proceed with sparse collection of PK and PD variables. These data can further support registration, may provide explanations for clinical trial success or failure, and are inexpensive to obtain when compared with the cost of enrolling patients.

1.4.2.4   Phase 4

Phase 4 studies are sometimes required by regulatory agencies. This can happen if the regulatory agency is interested in further characterizing safety, exploring new treatment indications, broadening label claims, exploring new drug combinations, or examining dosing in some special subpopulations (e.g., pediatric patients).

1.5   PHARMACOMETRICS AND REGULATORY AGENCIES

The FDA has promoted the role of pharmacometrics in the drug approval process through its approach to review of applications and by publishing its "guidances." The FDA has gained expertise in pharmacometrics from self-training within and by recruitment of new highly skilled personnel. The value of pharmacometrics continues to be evaluated at the FDA.

1.6   SUMMARY

Pharmacometrics is playing a major role in improving drug development and therapeutics. Improvements in drug development must come through creating and using novel pathways to approval and application of sound scientific principles, partly by applying mechanistic PM models. It is difficult to imagine a more efficient, powerful, and informative drug development process without the expansion of the role of pharmacometrics.

Pharmacotherapy is also in great need of improved dosing strategy selection for the avoidance of adverse events and the improvement in efficacy. This will come through the development of pragmatic PM models that provide knowledge

about drug behavior and how the drug can be optimally used. As more pragmatic PM models are developed, optimal dosing strategies can be implemented. The acceptance of pharmacometrics in drug use and development cannot, therefore, be overemphasized.

REFERENCES

1. J. A. Dimasi, Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **69**:297–301 (2001).
2. C. Connolly, Price tag for a new drug: $802 million. Findings of Tufts University Study are disputed by several watch dog groups. *Washington Post*, 1 December 2001.
3. Department of Health and Human Services, *Challenge and Opportunity on the Critical Path to New Products*. US Food and Drug Administration, Rockville, MD, 2004.
4. L. J. Lesko, M. Rowland, C. C. Peck, and T. F. Blaschke, Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *Pharm Res* **17**:1335–1344 (2000).
5. C. C. Peck, Drug development: improving the process. *Food Drug Law J* **52**:163–167 (1997).
6. P. J. Williams, A. Desai, and E. I. Ette, in *The Role of Pharmacometrics in Cardiovascular Drug Development*, M. K. Pugsley (Ed.). Humana Press, Totowa, NJ, 2003, pp. 365–387.
7. E. I. Ette, P. J. Williams, E. Fadiran, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).
8. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).
9. E. I. Ette, Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial. *Ann Pharmacother* **32**:818–828 (1998).
10. P. J. Williams, E. I. Ette, and C. J. Godfrey, in *Pharmacokinetics in Drug Development*, P. Bonate and D. R. Howard (Eds.). AAPS Press, New York, 2004, pp. 163—202.
11. A. Buchanan, Physiologic effects of the inhalation of ether. *London Med Gaz* **39**:715–717 (1847).
12. F. H. Dost, *Der Blutspiegel-Kinetic der Konzentrationsablaufe in der Krieslauffussigkeit*. G. Theime, Leipzig, 1953, p. 244.
13. E. Nelson, Kinetics of drug absorption, distribution, metabolism and excretion. *J Pharm Sci* **50**:181–182 (1961).
14. L. Michaelis and M. L. Menten, Der Kinetic der invertinwirking. *Biochem Z* **49**:333–369 (1913).
15. E. Widmark and J. Tandberg, Uber die bedingungen fur die akkumulation indifferenter narkoliken theoretische bereckerunger. *Biochem Z* **147**:358–369 (1924).
16. H. W. Haggard, The absorption, distribution, and elimination of ethyl ether III. The relation of the concentration of ether or any similar volatile substance in the central nervous system to the concentrations in arterial blood, and the buffer action of the body. *J Biol Chem* **59**:771–781 (1924).
17. R. Dominguez and E. Pomerene, Studies of renal excretion of creatinine I. On the functional relation between the rate of output and the concentration in plasma. *J Biol Chem* **104**:449–471 (1934).

18. T. Toerell, Kinetics of distribution of substances administered to the body I. The extravascular modes of administration. *Arch Int Pharmacodyn Ther* **57**:205–255 (1937).

19. B. L. Oser, D. Melnick, and M. Hochberg, Physiological availability of the vitamins: study of methods for determining availability in pharmaceutical products. *Ind Eng Chem Analyt Edn* **17**:401–411 (1945).

20. C. Lapp, Modes d'elimination des poisons. Definition de nouveaux coefficients toxicologiques. *Ann Pharm Fr* **6**:150–165 (1948).

21. W. Perl, A method for curve-fitting by exponential functions. *Int J Appl Radiat Isotopes* **8**:211–222 (1960).

22. E. R. Garrett, R. C. Thomas, D. P. Wallach, and C. D. Always, Psicofuranine: kinetics and mechanisms in vivo with the application of the analog computer. *J Pharmacol Exp Ther* **130**:106–118 (1960).

23. P. G. Wiegand and J. D. Taylor, Kinetics of plasma drug levels after sustained release dosage. *Biochem Pharmacol* **3**:256–263 (1960).

24. J. G. Wagner and E. Nelson, Percent absorbed time plots derived from blood level and/or urinary excretion data. *J Pharm Sci* **52**:610–611 (1963).

25. M. Gibaldi and G. Levy, Pharmacokinetics in clinical practice 1. Concepts. *JAMA* **235**:1864–1867 (1976).

26. J. R. Koup, J. J. Schentag, J. W. Vance, P. M. Kuritzky, D. R. Pyszczynski, and W. J. Jusko, System for clinical pharmacokinetic monitoring of theophylline therapy. *Am J Hosp Pharm* **33**:949–956 (1976).

27. R. Dungilson (Ed.), *A Dictionary of Medical Science*. Lea and Blanchard, Philadelphia, 1848, pp. 466, 648, 724.

28. H. Derendorf, L. J. Lesko, P. Chaikin, W. A. Colburn, P. Lee, R. Miller, R. Powell, G. Rhodes, D. Stanski, and J. Venitz, Pharmacokinetic/pharmacodynamic modeling in drug research and development. *J Clin Pharmacol* **40**:1399–1418 (2000).

29. N. H. G. Holford and L. B. Sheiner, Kinetics of pharmacologic response. *Pharmacol Ther* **16**:143–166 (1982).

30. N. L. Dyneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacologic responses. *J Pharmacokin Biopharm* **21**:457–478 (1993).

31. W. Kryzanski and W. J. Jusko, Mathematical formalism for the properties of four basic models of indirect pharmacodynamic responses. *J Pharmacokin Biopharm* **25**:107–123 (1997).

32. Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints: preferred definitions and conceptual frame work. *Clin Pharmacol Ther* **69**:89–95 (2001).

33. L. Zhang, S. L. Beal, and L. B. Sheiner, Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. *J Pharmacokinet Pharmacodyn* **30**:387–404 (2003).

34. P. J. Williams, J. R. Lane, C. C. Turkel, E. V. Capparelli, Z. Dziewanowska, and A. W. Fox, Dichloroacetate: population pharmacokinetics with a pharmacodynamic sequential link model. *J Clin Pharmacol* **41**:259–267 (2001).

35. M. Davidian and D. M. Giltinan, Some general estimation methods for nonlinear mixed effects models. *J Biopharm Stat* **3**:23–55 (1993).

36. M. Rowland and L. Benet, Pharmacometrics: a new journal section. *J Pharmacokinet Biopharm* **10**:349–350 (1982).

37. V. A. Bhattaram, B. P. Booth, R. P. Ramchandani, B. N. Beasley, Y. Wang, V. Tandon, J. Z. Duan, R. K. Bawaeja, P. J. Marroum, R. S. Uppoor, N. A. Rahman, C. G. Sahajwalla, J. R. Powell, M. U. Mehta, and J. V. S. Bobburu, Impact of pharmacometrics on drug approval and labeling decisions: a survey of 42 new drug applications. *Am Assoc Pharm Scientists J* **7**:E503–E512 (2005).

38. L. B. Sheiner, B. Rosenberg, and V. Marathe, Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm* **5**:445–479 (1977).

39. Department of Health and Human Services, *Guidance for Industry: Population Pharmacokinetics*. US Food and Drug Administration, Rockville, MD, 1999.

40. C. Maxwell, J. G. Domenet, and C. R. R. Joyce, Instant experience in clinical trials: a novel aid in teaching by simulation. *J Clin Pharmacol* **11**:323–331 (1971).

41. M. D. Hale, A. J. Nichols, R. E. S. Bullingham, R. H. Hene, A. Hoitsman, and J. P. Squifflet, The pharmacokinetic–pharmacodynamic relationship for mycophenolate mofetil in renal transplantation. *Clin Pharmacol Ther* **64**:672–683 (1998).

42. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

43. G. E. P. Box, Robustness in the strategy of scientific model building, in *Robustness in Statistics*, R. L. Laurer and G. N. Wilkinson (Eds.). Academic Press, New York, 1979, pp. 201–236.

44. P. Williams, J. Uchizono, and E. Ette, Pharmacometric knowledge based oncology drug development, in *Handbook of Anticancer Pharmacokinetics and Pharmacodynamics*, W. Figg and H. L. Mcleod (Eds.). Humana Press, New York, 2003, pp. 149–168.

45. E. Ette, P. J. Williams, J. R. Lane, Y. K. Kim, and E. V. Capparelli, The determination of population pharmacokinetic model appropriateness. *J Clin Pharmacol* **43**:2–15 (2003).

46. Department of Health and Human Services, *Guidance for Industry: Exposure–Response Relationships—Study Design, Data Analysis, and Regulatory Applications*. US Food and Drug Administration, Rockville, MD, 2003.

47. E. I. Ette, H. M. Chu, and C. J. Godfrey, Data supplementation: a pharmacokinetic pharmacodynamic knowledge creation approach for characterizing an unexplored region of the response surface. *Pharm Res* **22**:523–531 (2005).

48. E. I. Ette, V. Garg, and A. Jayaraj, Drug development a rational approach, in *Pharmacokinetics in Drug Development: Regulatory and Development Paradigms*, P. L. Bonate and D. R. Howard (Eds.). AAPS Press, New York, 2004, pp. 3–36.

49. R. Miller, W. Ewy, B. W. Corrigan, D. Ouellet, D. Herman, K. G. Kowalski, P. Lockwood, J. R. Koup, S. Donevan, A. El-Kattan, C. S. W. Li, J. L. Werth, D. E. Feltner, and R. L. Lalonde, How modeling and simulation have enhanced decision making in new drug development. *J Pharmacokinet Pharmacodyn* **32**:185–197 (2005).

50. J. Cross, H. Lee, A. Westelinck, J. Nelson, C. Grudzinskas, and C. Peck, Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999. *Pharmacoepidemiol Drug Saf* **11**:439–446 (2002).

51. V. F. Cosson, E. Fuseau, C. Efthymiopoulos, and A. Bye, Mixed effect modeling of sumatriptan pharmacokinetics during drug development. I: Interspecies allometric scaling. *J Pharmacokinet Biopharm* **25**:149–167 (1997).

52. M. Ekblom, M. Hammarlund-Udenaes, and L. Paalzow, Modeling of tolerance development and rebound effect during different intravenous administrations of morphine to rats. *J Pharmacol Exp Ther* **266**:144–252 (1993).

53. D. M. Ouellet and G. M. Plack, Pharmacodynamics and tolerance development during multiple intravenous bolus morphine administration in rats. *J Pharmacol Exp Ther* **281**:713–720 (1997).

54. L. J. Benincosa, F. S. Chow, L. P. Tobia, D. C. Kwok, C. B. Davis, and W. J. Jusko, Pharmacokinetics and pharmacodynamics of humanized monoclonal antibody of factor IX in cynomolgus monkeys. *J Pharmacol Exp Ther* **292**:810–816 (2000).

55. P. H. van der Graaf, E. A. van Schaick, R. A. Math-Ot, A. P. Ijerman, and M. Danhof, Mechanism-based pharmacokinetic–pharmacodynamic modeling of the effects of N6-cyclopentylanenosine analogs on heart rate: estimation of in vivo operational affinity and efficacy in adenosine A1 recpetors. *J Exp Ther* **283**:809–816 (1997).

56. K. S. Blesch, R. Gieschke, Y. Tsukamoto, B. G. Reigner, H. U. Burger, and J.-L. Steimer, Clinical pharmacokinetic/pharmacodynamic and physiologically based pharmacokinetic modeling in new drug development: the capecitabine experience. *Invest New Drugs* **21**:195–223 (2003).

57. J. M. Collins, D. K. Greishaber, and B. A. Chabner, Pharmacologically guided Phase 1 clinical trials based upon preclinical development. *J Natl Cancer Inst* **82**:1321–1326 (1990).

58. J. M. Collins, Pharmacology and drug development. *J Natl Cancer Inst* **80**:790–792 (1988).

59. F. Rombout, Good pharmacokinetic practice (GPP) and logistics: a continuing challenge, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, and U. A. Gundert-Remy, et al. (Eds.). Office for Official Publications of the European Communities, Luxemborg, 1997, pp. 183–193.

60. T. H. Grasela, E. J. Antal, and J. Fiedler-Kelley, An automated drug concentration screening and quality assurance program for clinical trials. *Drug Information J* **33**:273–279 (1999).

GENERAL PRINCIPLES

General Principles of Programming: Computer and Statistical

SASTRY S. ISUKAPALLI and AMIT ROY

2.1 INTRODUCTION

Although pharmacometricians are often involved in the development, modification, and use of computer code and programs, formal training in these skills is often neglected. Computer programming skills are acquired in an ad hoc approach, in which the minimal necessary knowledge to devise and code an algorithm is gained to solve the scientific problem at hand. This is not unexpected, as the scientific problem is of primary interest, and programming is simply a means to an end.

While the ad hoc approach to acquiring the necessary programming skills may have been adequate in the past, the need for sophistication in computer programming is increasing along with the complexity of computational problems being addressed by pharmacometricians. The programming approach that may appear to be expedient is often not the most efficient with respect to overall productivity. Additional effort in the initial stages of a project can save time and improve accuracy and overall quality of code in subsequent stages.

Although there are usually multiple ways in which a scientific programming problem can be addressed, adhering to standard programming approaches is an important step in development of high-quality programs. Standardization facilitates consistency and faster code reviews, and, more importantly, it helps a reviewer identify commonly occurring mistakes.

The aim of this chapter is to provide an overview of generally applicable good programming practices that could benefit pharmacometricians with regard to improving the quality and transparency of code, as well as increasing overall productivity. A set of techniques and practices is provided here that will be useful in writing better computer programs. The involvement of pharmacometricians with programming ranges from relatively simple code to complex, software development projects. Likewise, programming skills of pharmacometricians range from

novice to proficient, with proficiency usually gained from long experience. Rather than attempt to cover all aspects of programming in detail, this chapter covers the basics of writing good code and provides the reader with references to additional resources that provide more detail on other aspects of programming and software development.

2.2 PHARMACOMETRIC PROGRAMMING TASKS

Change is a dominant factor in scientific programming; hence, a scientific program needs to be easily readable and easily modifiable. In this sense, a scientific computer program is analogous to a scientific document, in that it should provide context, be readable, and contain appropriate references. Furthermore, a well designed program will often be useful far beyond what the original programmer intended, because it will be easily readable, modifiable, and expandable. There is extensive literature on basic programming techniques for scientists and engineers, but a majority of the literature focusing on programming practices is over three decades old (1–4). Many modern books dealing with programming are often focused on highlighting the features of a language, or advanced techniques involving specific programming platforms or approaches. Recently, there has been increased attention on good practices in software design (5, 6).

Pharmacometricians are often involved in programming tasks that span a wide range of complexity, ranging from writing a few lines of code to writing scripts and programs. These programming tasks can be classified according to a variety of attributes as shown in Table 2.1. Moreover, pharmacometricians may also be the domain experts on a software development team, providing guidance or input to other programmers. Therefore, much of the programming tasks demanded of a pharmacometric scientist involve writing not full programs from scratch but customizations of existing code or minor modifications to existing modules in order to create a program.

One example of systems where a model can be developed without much programming is ADAPT II (7), which provides templates of Fortran subroutines. In ADAPT II, the scientist is required to specify the model by adding code to existing templates of subroutines, in order to create a complete program. These subroutines can then be compiled and linked to other compiled code (object files) to create a stand-alone executable. Another example is the specification of models in NONMEM (8, 9). In NONMEM, the model is specified by a control file, which is then processed, to produce Fortran code that is compiled and linked to other object files to create an executable file. Although sophisticated programming skills are not necessary to develop models using these programs, some of the concepts described in this chapter will be useful in scripting even these relatively simple programs.

More extensive programming is often required in writing scripts or programs for software packages such as S-Plus (10, 11) or Matlab (12, 13). These two modern software packages are increasingly used by pharmacometricians: Matlab as a programming environment for numerical simulations and S-Plus as a programming environment for statistical data analysis.

It must be noted that there is a considerable overlap between the roles of these two packages, and both provide strong graphical capabilities. Although the princi-

TABLE 2.1   Examples of Different Types of Classifications Found in the Scientific Programming Space[a]

Programming experience	Novice to professional programmer
Scientific experience	Key scientist to programming support staff
Programming role	Use/apply others' code, review code, develop new software modules
Problem/model complexity	Linear models, algebraic equations, ordinary and partial differential equations
Randomness	Deterministic models, simple error models, stochastic systems
Software project complexity	Individual, local group, distributed group, production versus prototype versions
Complexity of the tools	Spreadsheet based, predefined modules (e.g., NONMEM)
Programming approach	Procedural, object-oriented, visual, symbolic, pipeline-based, event-based
Program dependencies	External databases, external web services, other programs; used as module in other programs
Program interfaces	Command line, noninteractive, distributed, web-based, embedded into other programs (spreadsheets)
Documentation complexity	Simple commenting/memos, detailed documentation published as reports
Quality assurance level	Error checks, automated tests, reproducing results, internal/external review
Extensibility and modularity	Single run models versus multiple run, cluster-based simulations

[a]Though the space spans a wide range, the general programming principles are applicable throughout.

ples of good programming practice described in this chapter are generally applicable for a variety of programming environments, they will be mainly illustrated using examples of Matlab code. Although it is possible to use much of the functionality of Matlab through the graphical user interface (GUI) or interactive commands, the full features of these systems can be utilized only through scripts. Furthermore, there are many advantages to writing scripts. First, scripts provide a record of the commands executed and facilitate the reproducibility of the results. Second, scripts provide a means for automating repetitive tasks and relieve the tedium and errors that commonly occur in performing repetitive tasks with a GUI interface, especially for computationally intensive tasks that have long waiting times between user input steps. Third, once a set of scripts that accomplish common tasks have been developed for a given project, they can often be modified for subsequent projects with a much smaller time investment.

Some pharmacometricians may be involved in complex software projects, such as the development of software for ADMET (absorption, distribution, metabolism, excretion, and toxicity) predictions or software tools that can be used by other scientists. Examples of such tools include Perl-speaks-NONMEM (14) or Xpose (15). Such tasks often require a diverse set of programming skills and strong programming practices.

2.3   OVERVIEW OF SCIENTIFIC PROGRAMMING METHODOLOGY

The programming paradigms applicable to scientific programming have often followed the developments in the field of software engineering. Some of the major paradigms applicable for scientific programming are briefly described in Table 2.2.

2.3.1   Scientific Program Development Steps and General Guidelines

Introductory programming books often provide resources for learning a programming language and programming syntax, for utilizing the development environment, and for compilation, execution, debugging, and optimizing of programs. All these techniques are directly applicable to scientific programming. Furthermore, there are a few additional points that a scientific programmer has to be aware of: (a) change is the dominant factor in scientific programming; (b) quality assurance is more important in scientific programming than in regular programming because it is often difficult to distinguish program errors or bugs from bad science; and (c) it is often very difficult to notice errors in the results.

A scientific program may start as a script for solving a specific problem and may find use in related areas. Sometimes, the program finds use in a much broader context. Some of the uses of the program can be (a) as one step in a sequence of steps involving multiple programs (i.e., in the form of a "pipeline"), (b) as a script that is invoked by another script, (c) as a function that is invoked by other functions or scripts, (d) as a program wrapped around a Monte Carlo type simulation or a parameter estimation module, (e) as a module wrapped around a graphical user

TABLE 2.2   Overview of Some of the Main Programming Paradigms and Approaches Applicable to Scientific Programming

Procedural programming	Modules or procedures are used as computational blocks
Flow-driven programming	Execution of code follows a well defined order
Event-driven programming	Execution of code depends on the events such as user clicks
Object-oriented programming	Objects, interfaces, and methods are used as computational blocks
Design patterns-based programming	Utilizing standard solutions to software design problems
Symbolic programming	Calculations are performed in a symbolic manner (e.g., Maple)
Visual programming	Assembling of "blocks" visually to form full programs
Pipeline programming	Output of one program is used as input of another (pipeline)
Collaborative programming	Deals with advantages (and issues) of multiprogrammer projects
Parallel/distributed programming	Deals with utilization of multiple machines
Web-based programming	Programming focusing on web-based interfaces

interface, (f) as a web-enabled program, and (g) as a program that is run multiple times on a distributed machine cluster. Therefore, it is prudent to follow good programming practices for all levels of programming tasks.

Schneider (16) recommends that a beginning programmer should concentrate on semantics and program characteristics of a programming language, and not just on the syntax. The concerns for programming style should be cultivated from the very beginning, and care must be taken to avoid the common mistake of initially writing beginning programs quickly with the idea of coming back later and then refining them. This prevents bad coding habits from ever developing. The programmer should also become familiar with and follow formal processes for debugging, program testing, and verification, as well as for documentation. Seeley (17) argues that following programming practices is more productive than simply using the latest tools.

Computer programming tasks in recent times have evolved from writing new code and modules to correctly linking existing modules. The majority of effort involved in solving a scientific programming problem is in identifying the appropriate design for the solution, and in identifying relevant existing modules; the linking of the modules becomes a simple task once the design is completed.

The following set of objectives with respect to the quality of scientific programs is recommended in this chapter: (a) program correctness, (b) reproducibility of results, (c) program readability (critical for code reviews), (d) maintainability (bug fixing and minor changes to the program), (e) ease of configuration change (e.g., parameter values and the constants used in the program), (f) portability and extensibility (ability to run the program on different systems and ability to link the program with other programs), and (g) performance (speed and disk space requirements).

The general steps involved in the development of a scientific program are common to programming tasks across a wide range of scales, from simple programs developed by an individual to complex software development involving a large group. However, implementation of individual steps varies depending on the type of problem solved, the scale of the project, and the level of quality testing.

These main steps are:

1. *Mathematical Formulation of the Scientific Problem.* In this step the scientific problem is formulated in mathematical terms and may involve reviewing the literature, identifying the appropriate mathematical model, and identifying sources for model parameter values.

2. *Algorithm Design.* Here, the problem has to be addressed from a computational framework viewpoint. Issues such as selection of a model solution scheme (e.g., choice of a differential equation solver, choices of appropriate modules for random number generation, etc.) are addressed at this stage.

3. *Design and Documentation of the Computer Code.* Here, the program is designed in a top–down approach. Interactions between the main program and individual modules are defined at this stage, along with brief documentation of the functionality of each module. At this stage, the program does not have much code—only definitions of the functions and parameters. The body of the functions is mostly empty at this stage. Changes to the program and the interactions between the modules can easily be made at this stage.

4. *Design of Test Cases.*   Representative test cases are identified and documented so that when the actual code is written, it can readily be tested. A brief review of the test cases is also done at this stage.

5. *Program Implementation.*   At this stage, a programmer can focus on individual modules. Typically, a programmer should develop simple, "unit tests" for individual modules at this stage. For modules with very few lines of code, these unit tests may be very simple, but in general writing unit tests is a good practice. As the complexity of the module increases (e.g., for the main module of a pharmacokinetic model), unit tests could involve calculating steady-state estimates with zero input doses (where many target tissue concentrations should reach zero) and very high input doses. Also, simple mass-balance tests can also be added at this stage. For example, when simulating systems involving multiple chemicals and reactions among them, an inert test chemical can also be introduced into the simulation and simple mass balances can be used for testing. The unit tests should be designed in a manner that facilitates easy debugging, so by definition they should be simple.

6. *Program Verification and Correction.*   At this stage, the programmer runs the code, fixes errors, and runs the test cases. If there are subtle errors, an interactive debugger can be used for stepping through the program. For complex simulations (e.g., those that run for several hours), programs can be monitored through log statements.

7. *Program Refinement and Optimization.*   At this stage, the program is refined and optimized. Feature enhancements, performance improvements, improvements in usability, and so on are common at this stage.

Adequate documentation and representative test cases are critical for developing good scientific programs. The documentation can be in the form of references (e.g., for assumptions used, mathematical models, and references for parameter values). Furthermore, when the programs are likely to be used by other scientists in a group, following a set of guidelines used in the group (or developing a set of guidelines if none exist) is a good step.

2.3.2   Tools for Numerical and Statistical Programming: Matlab, S-Plus, and Open Source Alternatives

The principles and practices discussed here are general in nature and are applicable to a wide range of scientific programming problems. They are also independent of the programming language and approach used. Specific examples are provided using Matlab, which is a programming environment for numerical simulations. These examples can also be readily applied to S-Plus, a widely used programming environment for statistical data analysis; however, it must be noted that there is a considerable overlap between the functionality of Matlab and S-Plus.

Matlab is a high-level scientific scripting language and an integrated development environment with interactive tools for visualization and several toolboxes addressing different computing areas such as statistics, database connectivity, and data mining. A pharmacometrician using Matlab may have to purchase Matlab toolboxes, such as the Statistics Toolbox, in addition to the basic Matlab license;

therefore, some individuals may find the cost of Matlab high. Fortunately, free, open source alternatives to Matlab exist: Octave (`www.octave.org`) is a high-level language, primarily intended for numerical computations, and Scilab (`www.scilab.org`) is a scientific software package for numerical computations. Both Octave and Scilab are similar to Matlab, and like Matlab, they both have large sets of toolboxes: Octave toolboxes are available in the form of the octave-forge package, while loosely coupled toolboxes are available for Scilab.

S-Plus is a statistical data programming language environment that follows the approach of programming with data. It is scalable and handles massive data sets and provides integrated tools for advanced analytics such as data mining. It also provides some advanced modules relevant to pharmacometricians, for example +SeqTrial™ for designing, monitoring, and analyzing clinical trials using group sequential methods. S-Plus license fees may also be an issue for some individuals. Free, open source alternatives to S-Plus include R (`www.r-project.org`) and Omega project (`www.omegahat.org`). R is very closely related to S-Plus, as both are based on the S software from Bell Labs; in fact, a majority of R code can run unchanged in S-Plus. A large set of modules for R are available at the Comprehensive R Archive Network (CRAN, which is part of the R Project).

The use of free, open source tools is suggested for pharmacometricians who may not have licenses for commercial software. However, the expenses associated with the licenses may not be significant for many organizations. A pharmacometrican can utilize the similarities between the proprietary and open source tools by developing the skills using the free tools and, if needed, transition to the proprietary versions later on.

One of the consequences of rapid advances in computer technology is that users are not constrained by the programming language or environment they use. In fact, many interfaces for invoking one language from another exist. For example, S-Plus can operate with SAS (`www.sas.com`) data sets. The Omega Project provides an R–Matlab interface (currently, an early release status) that facilitates a bidirectional interface between the R and Matlab languages that allows users of either language to invoke functions in the other language using the syntax of their choice. Matlab also provides interfaces to directly invoke functions in Fortran, C, C++, and Java.

2.3.3 Scientific Programming Resources

An overview of computational problem solving techniques for beginners can be found in Dijkstra (2) and Dromey (4). Several introductory textbooks on algorithm design are available freely (18–21). Textbooks based on specific programming environments and languages are useful in learning programming techniques—for example, for Matlab (12, 13, 22) or S-Plus (11, 23). Some of the books for advanced programming techniques are also freely available, for example, for object-oriented design (24), parallel computing (25, 26), user interface design (27), and agile-development (28).

One of the best ways to learn good programming skills is to read code from experts in the field. Often, reading and understanding code from an experienced programmer within an organization is also recommended, because it provides the novice programmer familiarity with the coding styles and approaches used in the

organization. Some of the approaches for beginning programmers include (a) repro-ducing the results from a working program, as this involves becoming familiar with the inputs and outputs used, getting familiar with the operating system and the pro-gramming environment, executing the program, and optionally postprocessing of the program outputs; (b) studying the code using "code browsers"; and (c) running the program in an interactive debugger and stepping through the code.

Programming productivity can be substantially increased by utilizing avail-able toolkits, libraries, development environments, and relevant programming approaches (29). Some of the productivity-improving features are available in inte-grated systems such as Matlab and S-Plus. For other features, or for programming in other languages, a programmer can use either specialized integrated development environments (IDEs) or general purpose toolkits. Though a detailed discussion of the available toolkits is beyond the scope of this chapter, some of the widely used general purpose tools include text editors such as XEmacs (`www.xemacs.org`) and ViM (`www.vim.org`); general purpose IDEs such as Eclipse (`www.eclipse.org`); debugging tools such as the GNU Debugger, gdb (`www.gnu.org/software/gdb`); code profiling tools such as the GNU Profiler, gprof (`www.gnu.org/software/binutils`); code browsing and publishing tools such as Glimmer (`glimmer.source-forge.net`); version control systems such as Concurrent Versions System (CVS) (`www.nongnu.org/cvs`); (30); and defect tracking systems such as bugzilla (`www.bugzilla.org`). Likewise, there is an large set of available libraries for general purpose scientific and statistical programming (31–34). Language-dependent librar-ies also exist, for example, Matlab libraries (35, 36) and S-Plus/R code (37).

A scientific programmer must be aware of available libraries and toolkits and must be familiar with the general tools and approaches for effective computer programming. An awareness of these tools and approaches will help in pursuing the corresponding features in the programming language of choice. For example, a Matlab programmer familiar with the notion of code browsing can either use the general purpose tool Glimmer (`glimmer.sourceforge.net`) or search for the feature in the Matlab Repository (35) and arrive at the Matlab code browsing tool, M2HTML (38). An awareness of features and utilities one can realistically expect in a programming environment will enable a programmer to seek similar features, often successfully, even in totally new programming environments. As an example, the Matlab IDE provides a majority of such features, and in some cases an auxiliary tool may be needed.

1. Enhanced editing ability, consisting of syntax-based code coloring/highlight-ing and automatic completion of variable names, facilitates faster coding as well as early detection of simple syntax errors (e.g., unbalanced parentheses, quotes). Some editors and IDEs also support an "outline mode" for navigat-ing large blocks of code.
2. Code "beautifying" tools enhance code readability via automatic indentation and line wrapping, as well as format code from diverse sources in a consistent manner.
3. Code browsing tools provide effective navigation of large blocks of code span-ning multiple files and thus are valuable for reviewing or studying programs written by others. Often, the code browsing tools allow publishing of the code

in a hyperlinked format (typically as a set of HTML files), which can be then viewed through a regular browser. The M2HTML tool (38) provides this functionality for Matlab.

4. Interactive debugging environments allow for tracing code execution, inspecting variables, and arbitrarily setting breakpoints inside the program. These allow for rapid location of errors. The Matlab IDE provides both a visual debugger as well as a command line debugger via the "dbstop" command.

5. Code profiling tools provide a summary report on the code execution, including time spent in different blocks of code, thus helping in optimizing the code. The Matlab IDE provides a profiler tool as well as the "profile" command.

6. Tools for periodic saving of program state provide value by (a) allowing an interrupted program to restart from a prior valid state and (b) allowing the user to monitor program progress by using the intermediate outputs. This is especially useful in the context of computationally demanding simulations that may run for days to weeks, because errors can be detected early by analyzing the intermediate outputs, and erroneous model runs can be stopped. Likewise, computational time is not lost when a correct model run is interrupted due to unavoidable problems. The Matlab system provides a "save" command to save the entire workspace or a set of objects.

7. Revision control tools allow easier management of source code changes in a transparent and efficient manner. Using these tools, a programmer (or a group of programmers) can easily track code changes, obtain a summary of changes from one version to another, and revert to any version based on either a version number or a date. Matlab provides an interface to several version control systems, for example, via the "cvs" command for CVS (30) and the "sourcesafe" command for SourceSafe (39).

2.4 GOOD PROGRAMMING PRACTICES: BASIC SYNTAX, CODING CONVENTIONS, AND CONSTRUCTS

The practices listed here are applicable to all aspects of scientific programming, including small segments of code or complete scripts, as well as modules that form a large software project.

2.4.1 Use Meaningful Names for Program Variables

Giving meaningful names to program variables is one of the simplest ways of enhancing the readability of code. The names of variables in legacy code are often cryptic, because the length of variable names was constrained in older programming languages (a maximum of 8 characters is allowed in Fortran 77). These constraints have been removed for all practical purposes in most of the current programming languages (such as Fortran 90, C, Java, Matlab, and S-Plus), which allow variable names that can be as long as 32 or even 256 characters.

Very short names (such as "t" to indicate simulation time) are easy to type, but (a) they are not informative about the nature and context of the variable, and (b)

(A) Well Named and Formatted Code

```
% Initialize compartment concentrations to zero
FOR iCmpt = 1:N_COMPARTMENTS
  FOR jChem = 1:N_CHEMICALS
    % Initialize concentration of chemical j in compartment i
    conc_cmpt(iCmpt, jChem) = 0;
  END % End of jChem loop
END % End of iCmpt loop
```

(B) Poorly Named and Formatted Code

```
FOR i = 1:N
FOR j = 1:M
c(ii, jj) = 0;
END
END
```

FIGURE 2.1 Code Block 1—impact of variable naming and code formatting on program readability.

they are likely to be misinterpreted or inadvertently redefined in another part of the program. It is preferable to use a meaningful name such as "sim_time" (note that in this scenario, the variable name "time" may not be appropriate because it may conflict with a system command, a reserved word, or an inbuilt function). Likewise, very long variable names should also be avoided, because (a) they are tedious to type and can lead to inaccuracies and lengthy statements, and (b) it is difficult to distinguish between long variable names that differ by only a few characters at the trailing part.

Short variable names are convenient and appropriate to hold temporary or intermediate values, such as counters in conditional loops. However, it is recommended that meaningful names be used even for counters. The code in Code Block 1 (Figure 2.1) provides examples of descriptive names for constants, variables, and temporary counters in conditional loops. Although this is a trivial example, the benefits of using descriptive counter variable names increase as the number of statements and nesting levels in the conditional block increase.

2.4.2 Use Consistent Naming Conventions for Program Variables

Many modern programming languages are also case-sensitive, and this feature can be used to advantage in communicating the type and context of a variable name. A convention often followed by Matlab programmers is to use all uppercase names for program constants. The case of a variable name can also be used to distinguish the context of the variable (local versus global) and variable type (scalar or vector versus matrix). In statistical programming, case is often used to distinguish between data set and column/variable names, thereby improving program readability. Some generally used naming conventions are presented in Refs. 40–42.

Consistent naming of variables is important for understanding the context of the variable and for writing reusable code. Some commonly used naming conventions are:

1. Use all uppercase names for constants and global variables.
2. Prefix global variables with an identifier, for example, GL_NUM_ COMPARTMENTS.
3. Use readable variable names either via underscores (num_compartments) or via mixed case naming (NumCompartments).

A related issue is the definition and initialization of constants. A commonly followed approach is to place definitions of constants together at the top of a unit of code (script or a module), so that the constants can readily be located. This also enables the programmer to identify at a glance the constants that have been defined.

2.4.3 Follow Organizational Conventions for Code Formatting

Proper formatting of code, such as indentation, wrapping of long lines, and splitting long formulas into shorter formulas, significantly enhances the readability of code, similar to a well formatted document. It also makes it easy to comment out or delete blocks of code. Many programming environments and modern, general purpose text editors have features for "beautifying the code." This includes appropriate automatic indentations and line wrapping (e.g., a two-space indentation for each nested conditional block). A consistent format not only helps in the readability but also highlights potential problems (e.g., a spurious "END" statement will alter the indentation in a visible manner).

2.4.4 Provide and Maintain Informative Comments

Providing comments that explain the purpose and logic of blocks of code is one of the most simple and effective ways of improving program readability. The main variables used should be commented along with major processing blocks (e.g., comments of the type "initializing the system" or "calculating derivatives"). Likewise, the end of loop constructs ("for," "while," and "if-then-else" blocks) should have a short, informative comment that mentions the conditional block that is being ended. This facilitates readability, especially for code that has several nested conditional statements. The major exception is for loops that consist of just one or two statements inside.

In general, comments should provide the context of a statement or a block of statements (i.e., why something is done) instead of just a literal translation of the statements themselves. For example, while commenting a break statement, indicating both the innermost loop (e.g., "exiting i_comp compartment loop") and the significance of the statement (e.g., "convergence reached" or "completed all dose inputs") is more informative than simply stating "exiting compartment loop."

It is very important to ensure that the comments and code are always consistent, as wrong comments can cause more harm than no comments. However, it is often the case that a good comment turns into a bad one because of changes in the code without updates to the comment. A good practice is to review and update comments whenever the code is changed significantly.

2.4.5   Avoid Segments of Commented Out Code

When extensive changes are to be made to an existing program, some programmers often tend to comment out working code and add new code, with the idea that the changes can easily be reversed. However, the commented out code often ends up staying in the program long after the code changes are finalized. Furthermore, programmers may add additional comments to describe why the code block was commented. This can lead to even more lines of difficult to follow programs cluttered with noninformative lines of text. The preferred method of revising code is to employ version control utilities such as CVS (30), which enable programmers to keep track of changes while maintaining the coherence of the code. The use of version control is briefly mentioned in Section 2.10.4.

2.4.6   Provide Documentation and References Along with the Code

It is essential to provide references for additional documentation when the code requires extensive documentation (e.g., statements involving a complex formula). This could also be in the form of an electronic document provided along with the code. Relying solely on comments to provide details can lead to comments overshadowing the code. Furthermore, text comments are limited in the type of information they can convey. For example, the documentation of a pharmacokinetic model can include the model schematic (a graphic), along with the model equations (mathematical objects), and additional references; such a document is significantly more useful than large chunks of text-based comments in the code.

2.5   GOOD PROGRAMMING PRACTICES: RELEVANT MATHEMATICAL CONCEPTS

Some of the basic mathematical requisites for scientific programming include understanding of (a) rules of operator precedence, (b) machine precision, (c) equality and inequality issues, and (d) potential for overflow/underflow of numbers. Related concepts such as relative and absolute differences are also important for scientific programming.

2.5.1   Operator Precedence

Operator precedence deals with the order in which different operations in a mathematical expression are evaluated. For example, in most programming languages multiplication has a higher precedence than addition. Understanding operator precedence is especially important when writing complex mathematical expressions, because it is a source for subtle errors. Using parentheses for grouping terms is a good technique, as it improves readability as well as reduces potential errors that creep in due to operator precedence issues.

2.5.2   Machine Precision Issues

These issues arise due to limitations in machine representation of numbers and fractions in terms of a limited number of computer bits (e.g., a decimal fraction

```
 1.  x =     0.00000000000000000001; % x correctly set to 1E-20
 2.  y = 5 - 4. 9999999999999999; % y (1E-16) set to zero
 3.  w1 = (x == y);  % evaluates to zero (false); Potential Error
 4.  w2 = (x < y);  % evaluates to zero (false); Definite Error
 5.  w3 = (x > y);  % evaluates to one (true); Definite Error
 6.  w4 = y * 1E16; % evaluates to zero (actual value 1); Error
 7.  w5 = x/y; % evaluates to inf (actual value 1E-4); Error
 8.  w6 = y/y; % evaluates to NaN (actual value 1); Error
 9.  c1 = (abs(x-y) < eps); % true; Equal within precision limit
10.  (x < eps & y < eps); % true, so avoid inequality comparisons
```

FIGURE 2.2 Code Block 2—a Matlab example highlighting common machine precision issues encountered.

such as 1/10 cannot be represented adequately with a limited number of bits in binary format[1]). Therefore, very small errors ("round-off" errors) are introduced, and these can sometimes accumulate over the course of a long simulation. In some programming languages, the problem is exacerbated by the choice of the variable type: for example, in Fortran, a "real" number is less precise than a "double precision" number. Therefore, depending on the problem and the choice of the variable type, the numerical errors can vary significantly. In scientific programming, such errors are sometimes encountered in the solution of systems of differential equations that are solved by numerical integration over a large number of time steps, with the accuracy of the solution highly dependent on the integration time step size and the duration of the simulation.

2.5.3 Equality and Inequality Issues

These issues arise due to the limitations imposed by machine precision. Very often, two quantities that should be identical will not pass the equality test because of the different ways in which they are computed. Sometime inequalities are also impacted. As an example, the Matlab statements in Code Block 2 (Figure 2.2) will produce unexpected errors (the statements were tested on Matlab Version 7.01; interactive; default setting of single precision).

   When Statements 1 and 2 are used to calculate the values of two small numbers, "x" and "y", one of them ("y") is incorrectly rounded off to zero, whereas the smaller of the two ("x") still retains nonzero value. Thereafter, all subsequent comparisons of "x" and "y" lead to unexpected and incorrect results. For example, the results of comparison in Statements 4 and 5 are definitely wrong, whereas the comparison in Statement 3 may or may not lead to an incorrect conclusion. There are some techniques to overcome these issues. For example, Matlab provides a variable "eps" that indicates the smallest floating point increment possible in a given precision. Statements 9 and 10 use "eps" in the context of comparing the difference of two numbers, as well as in deciding whether inequality comparisons can be made, and produce more predictable results.

---

[1]For example, 0.1 represented in binary becomes 0.000110011001100**11**... with the 0011 recurring; in general, only fractional numbers that can be represented in the form p/q, where q is an integer power of 2, can be expressed exactly, with a finite number of bits.

2.5.4 Overflow/Underflow Problems

These problems can sometimes arise in numerical calculations due to limitations of machine representation of very large or very small numbers. Overflow errors occur when the number to be represented is larger than what the computer can handle; thus, the number gets assigned a value of "Inf" (infinity). Likewise, underflow errors occur when the number to be represented is smaller than what the computer can handle; thus, the number gets rounded to zero. These issues lead to functions returning the values of "NaN" (Not a Number, or invalid number), for example, when two large numbers (infinity) are subtracted, or when two very small numbers (zero) are divided by each other (see Statements 6 to 8). Such issues may appear pedantic, but in scientific programming, very small numbers often result due to the small time steps in numerical solution of differential equations and are sensitive to the choice of units used to represent different quantities in the simulation.

2.5.5 Absolute and Relative Differences

These need to be used appropriately when the convergence of numerical simulations is to be evaluated, for example, to estimate steady-state values or to estimate the quality of a numerical approximation. This often involves a combination of relative and absolute difference criteria. Absolute difference refers to the magnitude of the difference between two values, whereas relative difference deals with the ratio of the difference to the actual value. When the values to be compared are very small (but substantially more than the machine precision), absolute differences are recommended to judge convergence. Likewise, when the values are very large, relative differences are useful in evaluating convergence. It must be noted that there are several exceptions to these recommendations, and the choice of the criteria depends on the problem at hand. A scheme that employs both the absolute and relative error criteria will provide a more robust means for evaluating convergence.

2.6 GOOD PROGRAMMING PRACTICES: REDUCING PROGRAMMING ERRORS

2.6.1 Explicitly Check for Errors Such as Division by Zero

Many programming languages handle numerical exceptions such as division by zero or square root of a negative number by aborting the program execution. However, some modern languages such as Matlab allow for computation to proceed despite such errors (see discussion on NaN and Inf in Section 2.5.4). Depending on the simulation and the programming environment setting, the following scenarios are possible: (a) the program aborts execution with an error indicating file name and line number of the offending code; this is common in simple programs written in C or Java; (b) the program continues execution and some of the variables will have infinite or nonnumber values; this happens often in Matlab and Fortran; (c) the program suspends execution at the first instance of an exceptional situation; this is common in interactive debugging environments when global error checks are enforced (e.g., in Matlab, one can set the option to interrupt when a NaN or Inf is encountered using the command "dbstop if naninf"); (d) the module makes a log

entry for the error condition and skips to the next iteration of the function program; this is common when multiple simulations need to be performed in one program, and erroneous simulations can be identified from the log files.

2.6.2 Avoid False Robustness in the Programs

Some programs are designed to be robust despite minor errors in the inputs and program state. Examples include web browsers, which are designed to do the best possible job in spite of errors. Such an approach should be avoided in scientific programming, because the robustness of the program comes at a cost: correctness. In a scientific program, it is often advisable for the program to fail noticeably when spurious conditions are encountered. For example, in Code Block 3 (Figure 2.3), the code in lines 1–3 does not perform any error check, the code in lines 5–8 "compensates" for errors in another module, whereas the code in lines 10–14 alerts the user when there is an error in the program. Though the right approach for error handling and alerting is often dependent on the situation, code that alerts when spurious conditions are encountered is preferable, unless otherwise dictated by the situation.

One practice for easy error detection is initializing variables to NaNs at the time of definition (or when error conditions are encountered). At any stage of computation, one can see if there are any invalid computations that are performed. Another practice for reducing errors is through explicit checks of function arguments. This is very essential, for example, in web-based business programs, where fraud is of high concern. In scientific programs, accuracy is of high concern, especially when subtle errors can result in seemingly reasonable, but wrong answers.

2.6.3 Check for Unused Variables

In languages such as Java and C, it is not possible to use a variable without declaring a type for it. So, the Matlab example in Figure 2.4 would not run properly in those languages. In this code, there is an error on line 11 in Code Block 4, a

```
 1. % A code block with no error checking
 2. conc = calculate_tissue_concentration(param1, param2);
 3. return conc; % returns conc values without error checks
 4.
 5. % A falsely robust code block
 6. conc = calculate_tissue_concentration(param1, param2);
 7. if (conc < 0), conc = 0; end % fix negative concentrations
 8. return conc; % compensates for errors in another module
 9.
10. % A fragile, but more correct code block
11. conc = calculate_tissue_concentration(param1, param2);
12.   if (conc < 0), error('Negative concentration encountered');
    end
13. % Above check raises an alert when an error is encountered
14. return conc;
```

FIGURE 2.3 Code Block 3—the role of appropriate error checks in scientific programs.

```
1. function result = get_blood_air_PC(scenario)
2. % Returns Blood-Air Partition Coefficient based on scenario
3. % Inputs: scenario: 1 => default, 2 => updated
4. % Outputs: blood_air_PC (Partition Coefficient)
5. % References: J. Doe, J.Pharm 2004, X. Y. Doe, J. Pharm.,2005
6. % Author: An Employee, Organization, Inc.
7. PC_BLOOD_AIR_DEFAULT = 0.2;  % J. Doe, J. Pharm., 2004
8. PC_BLOOD_AIR_UPDATED = 0.22; % X. Zmith, J. Pharm., 2005
9. PC_blood_air = PC_BLOOD_AIR_DEFAULT;
10. if (scenario == 2)
11.    PC_blod_air = PC_BLOOD_AIR_UPDATED;
12. end
13. result = PC_blood_air; % return PC for blood compartment

% Output from mlint
mlint('get_blood_air_PC')
L 11 (C 3-13): The value assigned here to variable 'PC_blod_air' is never used
```

FIGURE 2.4 Code Block 4—subtle errors in function definitions that can be identified only via auxiliary tools.

misspelled variable name. Some programming languages such as Java and C will produce compilation errors with similar code (compilation errors in which the variable PC_blod_air is not declared). However, this set of statements is valid in languages such as Matlab, and, therefore, no error will be reported. As a consequence of the error, the program will always use the default value. Furthermore, if the error is minor (e.g., the values of default and updated partition coefficients differ very little), the error will become very difficult to track. Fortunately, there are tools to identify such errors. For example, Matlab has a command for comprehensive code checking called "mlint." It must be noted that mlint is ideal for analyzing function files and is not as effective with script files in dealing with unused variables, since the purpose of a script may be just to initialize a set of variables to be used by another script. Code Block 4 also shows the output of mlint used on the code in lines 1–13. Matlab also provides "lint" report generation on all the files in a folder through a GUI. The report for the entire folder can be saved as an html file and easily reviewed.

2.6.4 Use "Catch-All" Statement Blocks in Conditional Constructs

When using a conditional statement such as "if-then-else" or "switch" statements, it is prudent to use a "catch-all" statement that addresses the unhandled cases. This may not seem important in the beginning, but it increases the robustness of the program when the code is used for different cases, because the program alerts the user when such errors occur. For example, when the statements in Code Block 5 (Figure 2.5) are executed for the set of chemicals handled by the program, the partition coefficient is assigned properly. However, if the "otherwise" portion of the "switch" statement is not used, the program would have produced silent errors when the code is used for new chemicals, by leaving the partition coefficient set to

```
switch (chemical_name)
  case {'cl2', 'chcl3', 'chloroform', 'ccl4', 'tce'}
     pc = estimate_pc_voc (chemical);
     % partition coefficients of volatile organics
 case {'hg', 'cd', 'as', 'As'}
    pc = estimate_pc_voc (metal);  % partition coefficients of metals
otherwise
    error(['Chemical ' chemical_name ' is not yet supported']);
end

% The corresponding "if" statement will have several lines of
% equality comparisons of the type
if (chemical_name == 'cl2' | chemical_name == 'chcl3' | ...
    chemical_name == 'chloroform')
```

FIGURE 2.5 Code Block 5—use of the appropriate structured programming construct: "switch" versus "if."

an unknown value. By including a "catch-all" statement, a noticeable error will be triggered upfront whenever the code is used for chemicals it does not handle.

2.7 GOOD PROGRAMMING PRACTICES: BASICS OF SCRIPT AND PROGRAM DESIGN

2.7.1 Avoid Monolithic Blocks of Code

In many programming languages, it is possible to write large, monolithic blocks of code. However, it is cumbersome to maintain and debug large blocks of code that require scrolling through several screens to be viewed in their entirety. One commonly used alternative to writing large blocks of code is to split large code blocks into multiple files, with each file tested individually and linked as a sequence of command files (also known as "including the files"). Though programs written in this fashion may appear to be "modular," they are in fact similar to single files with one large block of code.

2.7.2 Write Modular Code

The main aspect of modular code is that changes in one module do not alter the behavior of other modules. However, when multiple script files are included in the same module, they share the same "name space" (i.e., they all can access the same set of program variables). Thus, a minor change in one file, such as assigning a value to a variable, can have unforeseen consequences in other files. However, when modular code (via subroutines and functions) is used, changes internal to the function will have no consequences upon other functions. As long as the function parameters and return values are consistent, significant changes to the internals of the functions can be made, without affecting other modules. This reduces the chances of subtle, intractable errors.

In order to write maintainable code, programmers should break larger pieces of code into shorter functions (subroutines and procedures in some languages) that are small enough to be understood easily, have a well defined set of input arguments, and return corresponding outputs. This approach, though obvious, needs to be emphasized as it makes isolated small pieces of code easier to understand without having to understand the whole program at once. Likewise, once the small functions are tested, they can be assumed to work unchanged despite changes in other, unrelated functions.

Modularization via functions enables code reuse and avoids repetition of code. This approach is also superior to that of cut-and-paste of existing code into new code. For example, if there is an existing pharmacokinetic model for Drug A, and a similar model is needed for Drug B, it is preferable to modularize common functions and rewrite just the components that need to be changed, such as the initialization of the parameters. Otherwise, the programmer will inherit an additional task of keeping the two programs synchronized. When the code is modularized, bug fixes or improvements to a module are instantly reflected in all the programs that use that module. Modularization can significantly increase productivity, because modules that are used frequently are likely to be tested more often and improved in terms of accuracy and performance.

When modularization is carried to an extreme, it can lead to overengineering and also unreadable code. Programmers should exercise their judgment in deciding what level of modularization is appropriate for a specific set of problems.

2.7.3 Utilize Existing Modules and Libraries to the Fullest

A corollary to using modular programming is the use of modules developed by others. Most modern computer programming languages sport a wide range of modules in the form of libraries or toolboxes for solving a variety of problems. The type of problems they solve varies in scope: from sorting, searching, solution of linear equations or differential equations, random number generation, and plotting, among many others. Thus, a significant amount of computer programming can benefit from the "component model of programming," where the problem is often posed as finding appropriate modules from an existing toolbox and linking them to solve a specific problem. Despite the availability of well tested standard modules, some programmers tend to write new code to solve standard problems: for example, a module to solve an ordinary differential equation or a module to generate random numbers. It is recommended to perform a simple search to identify any existing modules before embarking on writing new ones, thus avoiding the problem of "reinventing the wheel." The main exception to this practice is the situation where license restrictions or organization policies necessitate developing new code to solve a standard problem.

2.7.4 Use Structured Programming

Structured programming is an approach in which a program consists of subsections, each with a single point of entry. Structured programming facilitates a "top–down" approach to program design, whereby the large scale structure of a program is mapped out in terms of smaller operations, which can be independently imple-

mented and tested. Structural programming is achieved by using hierarchical conditional constructs, such as "if-then-else," "switch," "for," and "while" for creating conditional branches of execution. This approach shuns the indiscriminate use of "goto" statements, which allows program control to jump to any line in the code identified by a line number or label and can make it difficult to follow the logic of a program.

The "goto" statement is sometimes used to direct program control when a program exception or error occurs. Alternative constructs that can be used in structured programming include (a) the "return" statement, which returns control to the end of the current function; (b) the "break" statement, which terminates the inner most loop; and (c) the "continue" statement, which returns to the next iteration of the innermost loop.

2.7.5   Use Appropriate Structured Programming Constructs

The choice of the structured programming construct used should convey the logic involved in a given operation. This is important because most of the constructs can be expressed in terms of other constructs: for example, a "for" loop can be written as either a "while" or "unless" loop. Some of the guidelines for the appropriate constructs to use are as follows:

Use "for" construct when the number of loop iterations is known beforehand. Likewise, use "while" construct when the number of loop iterations is not known beforehand. Cases include reading data from a file or from user input line by line until the end is encountered. Though this can be achieved by using a "for" loop with a conditional "break" statement, the "while" statement conveys the logic clearly. Some special cases require using "do-while" (when the first statement has to be executed before the conditional).

Use "switch" construct instead of multiple, nested "if-then-else" statements, especially when all the conditionals are treated at the same level. The resulting code is usually easier to read and follow. However, when different types of conditionals are tested, multiple, nested "if-then-else" constructs are preferable. An example for using "switch" versus "if" statements is shown in Code Block 5 (Figure 2.5).

2.7.6   Use Data Structures Appropriate to the Problem Under Consideration

A programmer should select the appropriate program types that properly define the computational problem. For example, if the number of compartments is a user-defined construct (i.e., the program is designed for solving systems of equations for a multicompartment PBPK model), the number of compartments becomes a parameter. However, if a PBPK model is for a specific implementation, a constant should be used to describe the number of compartments. Likewise, depending on the situation, a matrix may be more appropriate than a set of one-dimensional arrays. Similar choices have to be made with respect to selecting complex data structures versus default types provided by the programming language.

2.8 GOOD PROGRAMMING PRACTICES: MODULAR CODE DESIGN FOR FUNCTIONS

2.8.1 Restrict Use of Global Variables

Global variables are variables that are active at all stages of the code, while the scope of local variables is restricted to the function in which they are defined. Global variables should be used with care to avoid inadvertently setting a value in one module of code that could affect computations in another module. Use local variables as far as possible. Global variables are most appropriately used to define constants that do not change during the execution of the program (e.g., molecular weight of a chemical). They are a convenient means of passing values through more than one level of module hierarchy. An example with Matlab code for a differential equation model is given in Code Block 6 (Figure 2.6), in which global variables are used to pass values to the derivative function (which is not called directly from the code block where the constants are defined). The constants could have been defined in the derivative code, but defining them earlier is more efficient as the statements in the derivative code are executed repeatedly.

2.8.2 Pass Information Through Function Parameters and Not Through Global Variables

One of the advantages of global variables is that they are accessible from all components of the program. This also means that keeping track of the global variables becomes very difficult. A change in the values of a global variable in one function may trigger a difficult to notice change in another function. Therefore, passing information via function parameters is much more robust than passing information

```
(i)
y = solveMyODE('f', x0, x1, y0, dt, eps1, 23);
% Above function call is not informative

(ii)
y = solveMyODE(struct('func', 'f', 'xinit', x0, 'xend', x1, ...
                      'yinit', y0, 'tstep', dt, ...
                      'relerror', eps1, 'ODEMethod', 23));

(iii)
config.function = 'f';
config.xinit = x0;
config.xend = x1;
config.tstep = dt;
config.relerror = eps1;
% ODE (Ordinary Differential Equation) solver option
config.ODEMethod = 23;
y = solveMyODE(config); % passing config object to ODE solver
config.ODEMethod = 13;  % change just the ODE solver
y = solveMyODE(config); % invocation with modified options
```

FIGURE 2.6 Code Block 6—passing large sets of function parameters through custom data structures.

through global variables. This approach also allows for easier definition of test cases, since the only changes in the function state will be caused by changes in function parameters.

The main exception to this practice is when the same information needs to be passed through a nested set of functions. In PBPK modeling in Matlab, such a scenario is often encountered when a main program invokes a subprogram that invokes a differential equation solver.

2.8.3   Avoid Too Few or Too Many Function Parameters

A function with too few parameters is usually less flexible. However, a function with too many parameters is a good candidate for further modularization into multiple functions. The extra inertia in having to provide a large set of parameters to invoke a function will lead to an underused function; often, a programmer will use a simpler alternative.

2.8.4   Write Functions Using a Variable Number of Arguments with Reasonable Defaults

Many programming languages support defining functions that operate with a variable number of arguments, with a common example being the "print" function. A function that accepts a variable number of arguments along with reasonable defaults can provide great flexibility and functionality. The function will be easy to invoke, because it does not require a large set of parameters; but at the same time, it will be flexible enough for advanced users of the function. Matlab provides the feature of variable number of function arguments and function outputs through the constructs "varargin" and "varargout." This feature is often encountered in Matlab in the solution of differential equations: a novice can use the solver with default options and still get a reasonable solution, whereas an expert can tune the function performance by providing advanced options. Of course, there is an additional overhead involved in writing functions that handle a variable number of parameters, including checking whether required parameters are provided, what optional parameters are provided, and what parameters need to be set to default values. However, the code for handling such tasks is similar from function to function, and a well designed, flexible function usually is worth the additional code required.

2.8.5   Use "Try-Catch" Type Exception Handling

One of the most powerful features of modern programming languages is exception handling. However, it is significantly underused in scientific programming. Blocks of code that use exception handling consist of two parts—the normal program (also called as the "try" block) and the errors/exception block (also called as the "catch" block). The main advantage is that the exceptional conditions can be handled in one location without "cluttering" the code for main program flow. The exceptions from the lower level functions (e.g., a square root function) can be propagated to the higher level invoking function, which can then handle the error condition appropriately. Thus, the writer of the lower level function need not focus on how to handle the error condition and can simply focus on alerting the caller function about the error conditions. This enhances the modularization of the code as well as allows more code reuse.

2.8.6   Use Custom Data Structures with a Hierarchical Design

Most modern programming languages support user-defined, complex data structures and objects, and that feature can be utilized in writing clearer code. For example, in Matlab, instead of using variables such as `PBPK_Human_chloroform_Vmax`, `PBPK_Human_chloroform_Km` and `PBPK_Human_chloroform_PC_blood_lung`, one can write compact, easily readable code, by defining the variables as constituents of a custom data structure, as follows: `pbpk.human.chloroform.vmax`, `pbpk.human.choloroform.Km` and `pbpk.chloroform.PC.blood_lung`. Therefore, the parameters can be used in the most appropriate manner depending on the context. For example, in case of a pharmacokinetic module for chloroform, the parameters can be passed as param = pbpk.human.choloroform.

Now, the parameters $V_{\max}$ and K_m, can be accessed as `param.`$V_{\max}$ and `param.Km`. Likewise, all the partition coefficients can be accessed as `param.PC`. This approach provides flexibility in parameter assignment and parameter passing and improves readability.

2.8.7   Use Informative, Custom Data Structures for Function Parameter Passing

A function that takes a structure that has informative field names is significantly more readable than a function that takes a large number of parameters. For example, in Code Block 6 (Figure 2.6), a function call of the form shown in (i) is typically used and is not informative. However, by using data structures for parameter passing, as in (ii), the context of the parameters becomes clearer. The parameter passing approach in (iii) is similar to that used in (ii), with the added advantage that the data structures for parameter passing can also be reused. Functions designed in such a manner can also be easily extended to include more parameters without requiring changes in the intermediate calling functions; that is, changes in a function invoked via intermediate functions will not impact the intermediate functions.

2.9   GOOD PROGRAMMING PRACTICES: WRITING EXTENSIBLE AND NONINTERACTIVE PROGRAMS

Often, a numerical model has to be run for different combinations of parameter values. Examples include performing a large number of Monte Carlo simulations with a model to estimate the range of uncertainties in model outputs or distributions of outputs for a study population. Likewise, parameter estimation techniques, such as the Bayesian Markov chain Monte Carlo (MCMC) (43, 44) technique, also involve running the full model with varying sets of parameters. Therefore, a program should be designed upfront in a manner that facilitates noninteractive (automated) runs. This aspect is critical in software testing (45, 46).

The notion of running programs in a noninteractive mode is common in the area of server-based computing using operating systems such as UNIX and Linux. In contrast, PC-based computing has been predominantly interactive. The advantage of server-based systems is that a user can connect to the server, submit one or more "jobs" for execution, monitor the progress of the simulations for a period of time, set the job status to "background," and disconnect from the server. In such systems,

several users can use one server simultaneously and do not need to stay connected for the duration of the simulations. In case the users need to maintain an interactive session, some advanced tools, such as the GNU screen utility (`www.gnu.org/software/screen`), provide the feature of a "virtual interactive session" that the users can disconnect from and reconnect to as needed. These tools can be contrasted with the current techniques in PC-based computing, such as remote desktop (`www.rdesktop.org`) or virtual network computing (VNC; `www.tightvnc.com`), where only one user can effectively be connected to the server at a given time, and there are no easy means for automating connections to multiple machines.

The advantage of noninteractive programs has become more pronounced with the advent of powerful but inexpensive computing clusters. Typically, a user has access to several tens to hundreds of machines in a cluster. Thus, the ability to run a program in a noninteractive or detached mode without continuous monitoring is very useful. Furthermore, since the user's main computer (typically a desktop computer) is not occupied with multiple connections to the server, one can submit large running jobs to the server without affecting the desktop machine.

2.9.1 Provide a Usable Interface to the Model

An intuitive user interface, either command-line or graphical, is an important factor affecting model usability. Some of the relevant aspects include providing appropriate user input prompts, warnings, and diagnostics, when erroneous conditions are encountered, and user input validation and correction (e.g., re-prompting the user when an input error such as entering a text string when a numerical input is expected). This is complementary to the ability to run the model in an automated, noninteractive mode. Ideally, a program should be designed to operate in both interactive and noninteractive modes. A common approach for designing such programs involves running the model in an interactive mode when no command-line parameters are provided, and running it in an automated mode when the required parameters are passed via command line or through an input file.

2.9.2 Write Programs with Standard Formats for Inputs and Outputs

When the model uses a standard format for model inputs and outputs, it becomes easily extensible in the form of a link in a long chain of models. It also makes it easier for writing scripts to generate reports or plots from model outputs, to aggregate multiple model runs and perform additional analysis, and even to run multiple simulations based on other resources (e.g., using a database of chemistry parameters as an input to the model). Traditionally, the input and output formats are quite variable, and often a programmer would decide on the format based on the flow of the model. Some of the widely used general purpose formats include CSV (comma separated values, supported by most spreadsheet software) and XML (eXtensible Markup Language; `www.xml.org`). An effective approach is to utilize object storage features of the programming environment. For example, both Matlab and S-Plus provide an option to directly save a set of variables into an object file. These objects can be retrieved later by simply loading the object file.

One of the advantages of using the programming environment features to save objects is that, in addition to model inputs and outputs, extensive "metadata"

related to the inputs and outputs of a model can also be saved and later retrieved. The metadata can include (a) time when a simulation is run, (b) configuration options, (c) machine and folder paths for the simulation, and (d) the script that was run to produce the outputs. It facilitates easier reproducibility of results.

2.9.3   Write Easily Relocatable Programs

Relocatable programs are programs that can be run in isolation, without affecting earlier model runs. This requires that locations of input, output, and configuration files not contain absolute folder paths. This is important when multiple runs of the model are performed—each model run can be performed in a different folder and model runs from one simulation will not affect other runs.

2.9.4   Provide Ability to Uniquely Identify Results from Multiple Model Runs

When a model is used for performing multiple simulations, the "management" of simulations becomes a major issue. At a minimum, the simulations should be set up such that the outputs from different model runs can easily be identified. The simulation setup should allow a subset of model runs to be repeated without needing to repeat all the model runs. This is often achieved by providing a separate, appropriately named folder for storing inputs, outputs, and configuration files for each model run, and aggregating the model runs at the end of the simulation. Such an approach allows the model runs to be performed on a distributed cluster of machines.

2.10   GOOD PRACTICES: RELEVANT SOFTWARE ENGINEERING CONCEPTS

2.10.1   Follow Appropriate Directory Structures

Standardized directory structures for source code, documentation files, final executables, configuration files, and model inputs/outputs allow tracking the software development process and also help in easily integrating multiple, independently developed modules. Standardized directory structures allow easy detection of conflicts in the names of functions, scripts, or configuration files.

2.10.2   Utilize Available Libraries and System Tools

The advantage of using existing tools and libraries is that the programmer need not actively maintain or refine them. The modern programming experience often involves taking advantage of a diverse set of libraries, programs, and tools. Most languages provide interfaces to link modules from other languages (e.g., Fortran/C, Matlab/C, Matlab/Java), which can be utilized to link modules written in practically any language. The overhead involved in understanding new libraries, tools, and language interfaces pays off very quickly. Sometimes, the simple approach of using multiple programs in a "pipeline" is also effective. A common example is where a program's output is used to automatically generate plots and

reports using preexisting templates. When a programmer has flexibility and initiative in using multiple tools, there is an increased chance of using the right set of tools for a given task, within common constraints such as cost, as well as organization guidelines. License issues also play a major role, since there may be different types of restrictions that arise when using commercial (redistribution issues, code confidentiality, etc.) as well as freely available code (which may contain clauses that affect the derived code).

2.10.3 Use Appropriate Module from a Library

One problem with the availability of a large set of "standard" libraries is that at times it is possible to use the wrong module for a given task. A common situation involves the solution of differential equations. Some systems of differential equations contain derivatives that vary over wide scales, and these are known as "stiff" systems of differential equations. Therefore, a stiff differential equation solver should be used in these cases; otherwise, substantial numerical errors or convergence problems will result.

2.10.4 Use Software Revision Control Tools

Revision control helps in identifying changes to documents or code by incrementing an associated number or letter code, termed the "revision level" or simply "revision." Most modern revision control systems such as CVS (`www.nongnu.org/cvs`) (30) provide facilities to track changes based on user, time, or version number. For a group project, such systems are very critical. Even for a single programmer, such systems are essential because they provide some means of being able to reproduce a set of source files that satisfied some set of conditions in the past. These systems are vastly superior to ad hoc approaches for document control such as manual backups of directories. A further advantage of a revision control system is that the programmer has the flexibility to experiment with code changes without having to worry about manually managing extensive changes.

2.10.5 Embed Simple Testing into the Model: Simple Mass-Balance Checks

Embedding simple error checks into the model improves its robustness and extensibility. This can be done either at the mathematical model development stage (e.g., incorporating mass-balance checks in a pharmacokinetic model) or at the software implementation level (e.g., incorporating alerts whenever a negative concentration or a negative flow rate is encountered). The mass-balance type checks are valuable in the sense that they can highlight errors in both the mathematical model as well as the software implementation. The overhead associated with incorporating such error checks is warranted because of the benefits provided.

2.10.6 Utilize Test Cases and Peer Review in Program Design

Proper code testing and peer review of code and test cases are critical for scientific programming (46–48). One of the recent advances in software design methodologies includes the approach of "extreme programming" (XP) (6, 49). The XP approach

advocates the notion of software design by contract, where test cases are designed first and the code is written later (50). A major emphasis is also placed on designing the code such that the costs associated with code changes are lowered.

Software testing ranges from testing of an individual module (unit testing) (51) to the established discipline of formal software testing (45, 46). For an individual programmer, Humphrey (5) presents an insight into most major aspects of software development and relevant practices, including test-driven program design and reviews of design, code, and test cases.

2.11 SUMMARY

This chapter provides an overview of generally applicable good programming practices relevant to pharmacometricians. The guidelines provided here can be useful in developing correct, robust, and easily maintainable and extensible programs. These guidelines are targeted toward novice and intermediate programmers and may also provide some relevant tips to experienced programmers. Although sophisticated programming skills are not necessary to develop many pharmacometric programs, the concepts described in this chapter can be useful in writing even relatively simple scripts and programs.

The main focus deals with aspects of basic coding style, design issues, and tools that can be quickly used in improving the programming process. The style aspects focus on readability and standardization, which facilitate effective code reviews; the design aspects focus on structured programming and modular function design; and the software engineering discussion focuses on test design and program extensibility. Overall, the development of scientific computer programs is addressed from the perspective of writing scientific documents: they should provide context, be readable, and contain appropriate references.

Another aspect addressed in this chapter is that computer programming tasks in recent times have evolved from writing new code and modules to correctly linking existing modules. Programming productivity can be increased substantially by utilizing available toolkits, libraries, development environments, and relevant programming approaches. Once a good design or approach is employed, and relevant existing modules are identified, the linking of the modules to solve a pharmacometric programming problem becomes a more straightforward task.

ACKNOWLEDGMENTS

REFERENCES

1. J. M. Yohe, An overview of programming practices. *ACM Comput Surv* **6**:221–245 (1974).
2. E. W. Dijkstra, *A Discipline of Programming.* Prentice-Hall, Englewood Cliffs, NJ, 1976.

3. B. W. Kernighan and P. J. Plauger, *The Elements of Programming Style.* McGraw-Hill, New York, 1978.

4. R. G. Dromey, *How to Solve It by Computer.* Prentice-Hall International, Englewood Cliffs, NJ, 1982.

5. W. S. Humphrey, *PSP: A Self-improvement Process for Software Engineers.* Addison-Wesley, Boston, 2005.

6. S. W. Ambler, *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process.* Wiley, Hoboken, NJ, 2002.

7. D. Z. D'Argenio and A. Schumitzky, *ADAPT II User's Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software.* Biomedical Simulations Resource (BMSR), Los Angeles, 1997.

8. M. de Hoog, R. C. Schoemaker, J. N. van den Anker, and A. A. Vinks, NONMEM and NPEM2 population modeling: a comparison using tobramycin data in neonates. *Ther Drug Monit* **24**:359–365 (2002).

9. B. Frame, R. Miller, and R. L. Lalonde, Evaluation of mixture modeling with count data using NONMEM. *Pharmacokinet Pharmacodynam* **30**:167–183 (2003).

10. S. P. Millard and A. Krause, *Applied Statistics in the Pharmaceutical Industry: With Case Studies Using S-Plus.* Springer, New York, 2001.

11. W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS.* Springer, New York, 1999.

12. S. J. Chapman, *MATLAB Programming for Engineers.* Brooks/Cole-Thomson Learning, Pacific Grove, CA, 2002.

13. S. R. Otto and J. P. Denier, *An Introduction to Programming and Numerical Methods in MATLAB.* Springer, New York, 2005.

14. L. Lindbom, J. Ribbing, and E. N. Jonsson, Perl-speaks-NONMEM (PsN)—a Perl module for NONMEM related programming. *Comput Methods Programs Biomed* **75**: 85–94 (2004).

15. E. N. Jonsson and M. O. Karlsson, Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* **58**:51–64 (1999).

16. G. M. Schneider, The introductory programming course in computer science: ten principles, in *Papers of the SIGCSE/CSA Technical Symposium on Computer Science Education.* ACM Press, Detroit, MI, 1978.

17. D. M. Seeley, Coding smart: people vs. tools. *Queue* **1**:33–40 (2003).

18. R. L. Read, *How to Be a Programmer: A Short, Comprehensive, and Personal Summary.* Available freely from `http://samizdat.mines.edu/howto/` (2002).

19. H. Abelson, G. J. Sussman, and J. Sussman, *Structure and Interpretation of Computer Programs.* MIT Press, Cambridge, MA, 1996.

20. M. Felleisen, *How to Design Programs: An Introduction to Programming and Computing.* MIT Press, Cambridge, MA, 2001.

21. A. Downey, *How to Think Like a Computer Scientist: Learning with Python.* Green Tea Press, Wellesley, MA, 2002.

22. W. Gander and J. Hrebícek, *Solving Problems in Scientific Computing Using Maple and MATLAB.* Springer, New York, 2004.

23. M. J. Crawley, *Statistical Computing: An Introduction to Data Analysis Using S-Plus.* Wiley, Hoboken, NJ, 2002.

24. O. M. Nierstrasz and D. C. Tsichritzis, *Object-Oriented Software Composition.* Prentice Hall, London, 1995.

25. I. Foster, *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering.* Addison-Wesley, Reading, MA, 1995.

26. W. H. Press, *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing. Volume 2 of FORTRAN Numerical Recipes.* Cambridge University Press, New York, 1996.

27. J. Spolsky, *User Interface Design for Programmers.* `http://joelonsoftware.com/uibook/chapters/Fog0000000057.html` (2006).

28. K. Ka Iok Tong, *Essential Skills for Agile Development.* `http://www.agileskills.org/` (2006).

29. M. Giorgio, Survey of existing programming aids. *SIGPLAN Not* **11**:38–41 (1976).

30. K. Fogel, *Open Source Development with CVS.* Coriolis Group Books, Scottsdale, AZ, 1999.

31. NIST, *Guide to Available Mathematical Software.* `http://gams.nist.gov/` (2005).

32. GNU, *GSL–GNU Scientific Library.* `http://www.gnu.org/software/gsl/` (2005).

33. StatLib, *StatLib—Data, Software and News from the Statistics Community.* `http://lib.stat.cmu.edu/` (2005).

34. Netlib, *Netlib—A Collection of Mathematical Software, Papers, and Databases.* `http://www.netlib.org/` (2005).

35. MathWorks, *The MATLAB Central File Exchange.* `http://www.mathworks.com/matlabcentral/fileexchange/` (2005).

36. OctaveForge, *GNU Octave Repository.* `http://octave.sourceforge.net/` (2005).

37. CRAN, *CRAN—The Comprehensive R Archive Network.* `http://cran.r-project.org/` (2005).

38. G. Flandin, *M2HTML: Documentation System for Matlab in HTML.* `http://www.artefact.tk/software/matlab/m2html/` (2005).

39. L. C. Whipple and T. Roche, *Essential SourceSafe.* Hentzenwerke Publishing, Whitefish Bay, WI, 2001.

40. Wikipedia, *Programming Style—Appropriate Variable Names.* `http://en.wikipedia.org/wiki/Coding_standard#Appropriate_variable_names` (2005).

41. MathWorks, *Programming (Version 7)*, 2005.

42. S. McConnell, *Code Complete.* Microsoft Press, Redmond, WA, 2004.

43. F. Jonsson and G. Johanson, The Bayesian population approach to physiological toxicokinetic–toxicodynamic models—an example using the MCSim software. *Toxicol Lett* **138**:143–150 (2003).

44. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* Chapman & Hall, Boca Raton, FL, 1998.

45. R. Patton, *Software Testing.* Sams, Indianapolis, IN, 2001.

46. C. Kaner, J. L. Falk, and H. Q. Nguyen, *Testing Computer Software.* Van Nostrand Reinhold, New York, 2002.

47. K. E. Wiegers, *Peer Reviews in Software: A Practical Guide.* Addison-Wesley, Boston, MA, 2002.

48. W. S. Humphrey, *A Discipline for Software Engineering.* Addison-Wesley, Reading, MA, 1995.

49. J. Newkirk and R. C. Martin, *Extreme Programming in Practice.* Addison-Wesley, Boston, MA, 2001.

50. K. Beck, *Test-Driven Development: By Example.* Addison-Wesley, Boston, MA, 2003.

51. P. Hamill, *Unit Test Frameworks.* O'Reilly, Sebastopol, CA, 2004.

Validation of Software for Pharmacometric Analysis

GARY L. WOLK

3.1 INTRODUCTION

The development, installation, and utilization of software for pharmacometric studies require the pharmacometrician to interact with at least three different organizational entities. Management must first be convinced of the need, and the appropriate expense must be justified, for implementing the tools regarded as necessary by the pharmacometrician to perform a successful analysis. Next, there is the interaction with the local suppliers of technology, the information technology (IT) group. This interaction is critical to determining the timeliness and the success of the implementation process. Finally (and perhaps most important) is the interaction of the pharmacometrician with the regulatory group responsible for the software validation process.

The responsibilities of the scientist will vary, depending on the organizational size. If the pharmacometrician is employed by a small or startup pharmaceutical or biotechnology firm, it is plausible that the pharmacometrician may be filling all three of these roles—clinical developer/manager, information technologist, and regulatory specialist. In this instance, interdepartmental delays become nonexistent, but the burden on the pharmacometrician is immense. For scientists working in medium size institutions, there is probably a specialist available from each area, but the burdens on each group tend to be immense since the company is more than likely in a "growth" mode. Finally, in a large corporate environment, the scientist is confronted by the possibility of dealing with a less personal, highly specialized IT or regulatory organization or, possibly, organizations that have been specifically devoted to business segments such as clinical development. In a sense, this last condition is the closing of the business organizational loop where one person is responsible for the entire process to a set of organizations that is entirely focused on the success of this particular part of the pharmaceutical realization process.

This chapter outlines the software implementation and validation process, to an extent that the pharmacometrician could in fact establish the quality assurance infrastructure, implement the hardware and software, and validate the implementation independent of dedicated internal IT or regulatory resources. Though such an approach is not recommended, the purpose of this chapter is to give the scientist a clear understanding of what is required in order to be successful in such an endeavor.

First, we review the concepts behind software quality assurance, testing, and validation. We review the process from the historical perspective of how other industries have faced these quality assurance issues, the role of independent organizations, and finally the role that federal regulatory agencies have played and how each of these has impacted the validation process in the pharmaceutical industry (1, 2). The rule-making efforts of the US FDA in the last 5 years, in particular, the 21CFR11 guidance (3, 4), is discussed in the context of this historical perspective. We also note the critical issues that face pharmacometricians in executing their scientific methodology: obtaining/finding data, creating/defining models in software, creating/finding results, and reproducing analysis.

We then outline the basic methodology for software validation: quality assurance practices (corporate policy, standard operating practices, validation processes), technology practices (assuring the proper infrastructure, influencing and participating in the IT process), and the process for making "buy or build" decisions. Often the decision is to buy and then build on to the software base. This is particularly true of software tools that allow the pharmacometrician to either automate existing software processes or design variations on existing algorithmic routines offered by the commercial tool.

The validation process is outlined from writing user requirements specification to testing and validating specific analysis using estimation methods. This is followed with brief examples of validation approaches for some commonly encountered software, such as S-Plus®, SAS®, WinNonlin®, and NONMEM®.

3.2   SOFTWARE DEVELOPMENT AND IMPLEMENTATION: BACKGROUND

In the late 1980s at AT&T Bell Laboratories, it came as quite a shock to be told that the "quality" of our work needed to be addressed. The scientific staff was insulted and the nontechnical managers who implemented "quality improvement" programs based on the Japanese models of the time were without a clue as to why there would be such resentment. It took several years for all to realize that, indeed, the *quality of business practices* that surrounded R&D efforts needed improvement, not necessarily the quality of the technical effort. The processes surrounding R&D—documentation of work, sharing of information, the need to avoid duplicate effort—soon were understood to be significant areas of improvement that both technical staff and nontechnical managers could work together on to improve the overall nature of the business. The manufacturing division, the former Western Electric, had indeed been a center of quality improvement and statistical process control 20 years earlier. Telephony transmission and switching systems, by their nature of being large, complex, engineered entities, had always been subject to high levels of review and quality assurance.

What remained was to take the appropriate pieces of the quality assurance world and implement them in such a way that scientists were able to work in a "structured" framework, while at the same time assuring that the creativity of the scientists was not stifled. In the ensuing years, many industries began to adapt the guiding principles of ISO 9000 (5), using the actual certification process as a way both to identify and improve business processes and to leverage certification as a marketing tool.

In parallel with these industrywide quality improvement efforts, the software industry had recognized the need to identify processes and standards that assured the quality of commercial software. ANSI (6) and IEEE (7) have been issuing practice standards and definitions for many years in an effort to unify quality assurance methodology in the software development industry. Furthermore, the software industry recognized early on that establishment of quality principles prior to the initiation of a development effort reduced the cost of repairing faulty software later in the process (8).

In the early 1990s, the implementation of the Clean Air Act, along with major changes to other environmental laws and regulations, produced a tremendous effort in the field of data acquisition and analysis, which clearly needed to be aided by advances in the information sciences. Hence the Environmental Protection Agency (EPA) issued guidance in late 1990, the EPA Good Automated Laboratory Practices (GALPs) (9), that were the first effort by a regulatory agency to assure the proper use of IT in the acquisition and analysis of regulated data. The Food and Drug Administration (FDA) had previously taken the position that most good laboratory practice (GLP) and good clinical practice (GCP) processes involving information systems were covered by existing regulation (10). Although a guide to inspection of computerized equipment in drug processing (11) and a technical reference (12) on software development activities were issued in the 1980s, the major FDA guidance on the use of electronic records and systems was not issued until 1999. Once that guidance, 21CFR11, was issued (13), an entire industry arose to attempt to explain, implement, and modify the guidance (7).

In general, there are certain basic quality assurance principles that can be invoked that will satisfy the spirit, if not the fine detail, of most regulatory requirements:

1. Document the processes used to generate, accept, analyze, store, and archive data and analytical results.
2. Document the physical and logical security of hardware and software systems used on regulated data.
3. Document the installation and testing of hardware and software used on regulated data.
4. Document that the system design achieves the intended purpose/use.
5. Document performance, both initial and ongoing, of the software system.
6. Document training and education backgrounds of the users and providers of the systems.
7. Document that the business practices are in place to operate, backup, and recover (including disaster recovery) regulated software systems.

Each of these issues focuses on documentation. The purpose of the validation process and the generation of process standards (or standard operating

procedures—SOPs) are to establish a documentation framework a priori, rather than de facto with regard to the installation and use of key software. All of the processes listed above occur within disciplined scientific organizations. The validation methodology is used to demonstrate the structure of these processes for the purpose of both internal and external review.

3.3 METHODS

A successful validation strategy is aided by several elements including:

1. A corporate policy on quality assurance/validation.
2. Existing, corporatewide SOP infrastructure and pharmacometric specific SOPs.
3. Definition of the validation process.
4. Understanding the user requirements generation process.
5. Identifying the system specification for a particular implementation.
6. Understanding the current information technology infrastructure/ organization.
7. Recognizing the constraints of "building" versus "buying."

We discuss each of these in turn.

3.3.1 Corporate Policy

In the case of industries that decide to pursue ISO 9000 certifications, the role of management is well defined (5). The standard clearly states that management will define and document its policy, objective, and commitment to quality. The burden of implementing, explaining, and maintaining the quality plan is clearly on corporate management. A similar approach needs to be undertaken in approaching validation of regulated systems. A clear corporate policy document should exist, which:

1. Establishes a working group to define and maintain policy and objectives regarding validation of software systems.
2. Ensures that employees are trained and retrained on the policy.
3. Empowers the resources necessary to carry out the policy.

In the absence of support from the highest levels of corporate management, it is unlikely that the competing priorities of clinical development, information technology, and regulatory affairs will somehow "align" to enable the success of a validation project.

3.3.2 Establishment of SOP Infrastructure

The first priority of a regulatory group should be the establishment of "SOP on SOPs"—that is, how they are to be created, reviewed, implemented, and revised. If a policy applies across corporate groups including information technology,

pharmacokinetics, data management, biostatistics, regulatory affairs, quality assurance, and materials management, then the pharmacometrician will have model SOPs as well as the collegial support necessary to implement a procedural structure that may be used to clarify workflow and serve as a training tool for new scientists.

In particular, the process by which preclinical and clinical pharmacokinetic/pharmacodynamic (PK/PD) data is received, identified, and analyzed (at least for initial parameters such as AUC and $\tau_{1/2}$) should be documented in a series of SOPs. Furthermore, the manner in which such data and analysis should be stored for latter retrieval is also a key consideration for optimal efficiency in the drug development process. The process by which PK summaries and reports are approved and released to other groups must also be documented in order to prevent misunderstandings. The procedure for use of randomization codes by the PK group must clearly be documented by an SOP, consistent with the needs and requirements of data management, biostatistics, and regulatory groups. Table 3.1 shows a plausible sample of SOPs that could be written to encompass the activities of both clinical and preclinical PK and PD analysis. It should be noted that many of the chapter titles in this text could also serve as the basis of clinical pharmacology SOPs!

Given the current desire of management to be able to "mine data" and "see trends across studies" and the availability of PK/PD repository systems (which we discuss more fully later), the first and foremost operating procedure requirement in pharmacometrics is the definition of key metadata that describes the process flow. Metadata, from the information science perspective, is simply information that describes data: that is, where the data goes, what the data is, and what possible

TABLE 3.1   Standard Operating Procedures for Clinical Pharmacology

SOP #	SOP Title
PKPD001	Training requirements for pharmacologists and toxicologists
PKPD002	Definition of nomenclature: project, study, indication, NCE ID, etc.
PKPD003	Review and approval process for PK summaries and related reports
PKPD004	Standards for PK data analysis: basic parameters to be obtained
PKPD005	Use of blinded data
PKPD006	PK analysis standards, data preparation, statistical analysis, expected output for clinical bioavailability studies
PKPD007	PK analysis standards, data preparation, statistical analysis, expected output for clinical bioequivalence studies
PKPD008	PK analysis standards, data preparation, statistical analysis, expected output for human dose proportionality studies
PKPD009	PK analysis standards, data preparation, statistical analysis, expected output for drug interaction studies
PKPD010	PK analysis standards, data preparation, statistical analysis, expected output for clinical renal studies
PKPD011	PK analysis standards, data preparation, statistical analysis, expected output for first dose in human studies
PKPD012	PK analysis standards, data preparation, statistical analysis, expected output for compartmental study types

value it may have. For example, having consistent definitions and values for the words "Portfolio," "Project," "Protocol," "Study," "Study Design," "Study Type," "Indication," "Method," "Period," "Phase," and "Relative Nominal Time" can lead to a dramatic increase in the ability to find and leverage information within clinical development. Unfortunately, there are enough clinical data management, repository, and laboratory information management systems (LIMSs) available to completely confuse the end user as to how the corporate metadata matches a software vendor's definition. Nevertheless, a group of scientists who have established procedural definitions of such metadata a priori have built a common ground that can serve as a basis for leveraged information management.

3.3.3 Definition of the Validation Process

In general, the validation process should also be defined by several SOPs, originating in either the regulatory or information technology groups. Table 3.2 shows a sample list of IT or Quality Assurance SOPs appropriate to the task. The validation process generally will consist of the following:

1. Validation Project Plan is a summary document identifying software, hardware, and related systems involved in a specific validation effort. The document explains the approach that will be employed, the responsible parties, and the expectations of those parties for each task involved in the validation.
2. User requirements specification is the responsibility of those end users who have identified the need for the system. It must adequately define the functional requirements of the system/software so that the end users can satisfy the stated business requirement.

TABLE 3.2   Standard Operating Procedures for Information Technology or Quality Assurance

SOP #	SOP Title
QA001	Format, functionality, and maintenance of standard operating procedures
QA002	Membership and purpose for the software validation standards committee
QA003	Validation process: validation planning, user requirements, and system specifications
QA004	Installation qualification protocol format and requirements for software
QA005	Operational qualification protocol format and requirements for software
QA006	Performance qualification protocol format and requirements for software
QA007	Change control procedures
QA008	Deviation procedures
QA009	Reporting "out-of-specification" events
IT001	Physical security procedures
IT002	Logical security procedures
IT003	Backup and recovery procedures
IT004	Hardware installation qualification procedures
IT005	Software development life cycle practices and procedures

3. Systems specifications provide all the information needed for the technical implementation of the system. This includes hardware, networking connections, and backup requirements as well as all information needed to install, operate, and qualify the performance of the system. It generally includes all of the qualification protocol documents (installation, operation, and performance qualifications) created during the validation process.

4. Completion and change control is the closure of the process and a methodology for maintenance activities.

3.3.4   Understanding the User Requirements Specification (URS) Process

This process is usually the first exposure of the pharmacometrician to the validation process. It is a difficult first step, where the scientist must document the use of a tool (which is of obvious utility from the perspective of the pharmacometrician) to an audience that may not have a good understanding of the clinical development process. The major point is that the process can be quite generic for many of the software tools utilized in PK/PD analysis. For example, the user requirements specification for implementing S-Plus, SAS, or Graphpad Prism® could all be essentially the same document. Similarly, NONMEM, WinNonlin, WinNonMix®, Kinetica®, and ADAPT II would contain the same basic set of user requirements. Specific capabilities that would be used for a particular software tool would need to be identified, but the basic form of the requirements is the same. Once again, having a sound basis set of SOPs that actually describes the acquisition, analysis, and reporting requirements for clinical data will enable the pharmacometrician to cross-reference the particular software capabilities with the technical (business) process.

For systems such as PK/PD repositories, a broader view is needed. Such systems by definition are intended to exchange data with other systems and integrate with analytical tools such as those described earlier. In this case, the pharmacometrician needs to have a well established process in place and be able to document how such a repository system will be implemented to either augment or replace current manual processes.

3.3.5   Identify the System Specification for a Particular Implementation

The selection of platforms (i.e., UNIX versus Microsoft Windows Server) is primarily within the realm of system specification rather than user specification. Nevertheless, it is useful for the end user to consider early on which tools are preferred and which platform will be used or whether several platforms might be utilized (depending on business requirements). We discuss this issue further when the interaction with the IT group is reviewed. The pharmacometrician must be able to specify key system requirements with regard to recovery of data and archiving. Furthermore, the end user needs to participate heavily in the definition of the operational qualification protocol, since it is this protocol that will determine if the software is meeting the basic user requirements that have been recorded in the URS. Finally, the performance qualification is the responsibility of the pharmacometrician, since this testing will determine whether the software system is functioning within the business/technical needs of the end user.

3.3.6   Information Technology Infrastructure and Organization

The interaction between clinical pharmacology and the IT resources, whether internal or externally by contract, is of paramount importance when considering the productivity and analytical capability of the pharmacometrician. Once the pharmacometrician has clearly stated her/his software needs, based on the URS, the actual definition of the overall systems that will be used to service the needs of the pharmacometrician must be decided.

The clinical pharmacology area is one that is subject to the same regulatory demands of other clinical areas such as biostatistics or clinical data entry and validation, yet it is a discipline that utilizes scientific methodologies that are closer in reality to discovery and preclinical drug development. That is, modeling of data, attempting to establish the validity of a hypothesis based on accumulated data and prior scientific knowledge, is the process employed. While some variables and covariates may be well defined and understood, in many cases, especially in population-dependent studies, it is the "expert system" of the scientist's experience and ability that unveils the critical issues surrounding pharmacokinetic, pharmacodynamic, or toxicity effects. To this extent, clinical pharmacology, while considered a development activity, is more closely akin to a discovery process. In general, drug discovery areas such as medicinal chemistry, target identification and structure, or preclinical assay development are not subject to the regulatory information system requirements that clinical pharmacology must follow. Hence, the IT support structures normally associated with areas such as clinical development, which more typically involve electronic document control or clinical database management, need to be imbued with a technical understanding of the work of the pharmacometrician. Ideally, organizations should strive to identify a pharmacometrician, or other members of the scientific staff, with an interest in IT. Such individuals would not be "lost" to PK/PD research but rather would become an invaluable asset in communicating the specific needs of clinical research. Nevertheless, the usual situation is one where a computer engineer needs to be educated as to the needs of the pharmacometrician. If, in fact, that engineer is not devoted entirely to the clinical area, the probability of a successful interaction will decrease dramatically. It is improbable that an IT individual can successfully support business software dealing with human resources, purchasing, and customer relationship management while at the same time understanding the needs of the pharmacometrician to create a model (perhaps by generating new code to do so), automate the analysis of a large number of studies, and then generate a PK report using completely separate tools.

The ultimate goal of the IT staff assigned to the clinical development groups must be customer service. In order to increase the throughput and accuracy of the pharmaceutical realization process, the clinical development area must be given the IT resources and attention necessary to determine the efficacy and safety of the subject chemical entity. If such resources are available, they must be encouraged to serve as advocates for the pharmacologists they support to IT management and corporate management. Once again, it is up to the pharmacometrician to establish a relationship with the IT support structure that encourages this attitude on the part of the IT support personnel. The clinical pharmacology group should at the least identify an individual within their organization as the liaison with IT. That individual should be included in meetings held within IT regarding policy,

infrastructure, and resources so that there is a clear source of information regarding IT infrastructure. In smaller company situations, where the pharmacometrician may in fact be the IT support person, clearly the ability to influence and participate in the IT support process is critical.

3.3.7   The Buy or Build (or Buy and Build) Decision

Most clinical organizations take for granted that their key capabilities lie in PK/PD rather than in software development. Nevertheless, some of the most utilitarian tools used in pharmacometrics have been written by pharmacologists. While many of these have arisen from university endeavors, several commercial packages began as "skunks works," projects that have evolved into private companies that provide valuable tools. The point is that it is probable that, within current organizations, there are individuals who are certain that a better mousetrap is within reach. Furthermore, most commercially available tools have specifically enabled programming and automation tools (such as WinNonlin) or interfaces (such as S-Plus) where custom development is not only possible but more than likely would have a positive impact on the drug development process. The issue to consider when going down the "home-brewed" or automation road is that the software development process, as discussed in Chapter 2, must be well documented *before* the development process begins. Just as in the documentation of the clinical drug evaluation processes with SOPs, the clinical group must become familiar with and document (via SOPs) the software development life cycle process (SDLC) as it will be implemented within the group; or possibly, if in a large company, it is plausible that the IT group already has established SOPs for SDLC. The starting point for commercial and internal development is exactly the same—the user requirements document. In addition, a functional requirements document should be written, outlining the details of how the specific functionality of the software (i.e., subroutines used, function of a drop-down menu, a panel of buttons, or what to type in as a command to execute some specific task) needs to be written. Unit test plans—how the person or persons generating the code will determine if well defined subunits of code are working—must also be generated. Finally, installation of the code (or the macros, if a commercial tool is being automated) and operational and performance qualification should be performed in the same manner as any commercial application would be. Another important consideration is that some type of source code control system must be identified and employed so that the history of the software development process, as well as any change control process after installation, may be documented.

Alternatively, if the decision is made to buy only commercially available software, or only commercially developed add-ons or automation scripts, then the pharmacometrician needs to participate in the key processes used to evaluate the vendor. The occurrence of key quality failures in widely used software has been previously documented (14). Therefore, the pharmacometrician should be intimately involved in the *vendor audit process*. If the vendor is not performing the quality assurance procedures just outlined for internal development, the cost (both in quality and accuracy of future work) will be in jeopardy. As discussed later in the section on validation documentation, the ability to state what the vendor's quality processes are will mitigate the need to perform functional software testing at the same level that has already been executed by the vendor's quality assurance group.

3.4 VALIDATION PROCESS

The primary issues surrounding the documentation of the validation process and the role of the pharmacologist are now addressed. The main areas of concern to the end user are:

1. What documentation needs to be created?
2. What is the order or priority of the document generation process?
3. Who is responsible for various sections of the documentation?
4. What are the content, format, and future (i.e., what are the maintenance requirements.) of the documents?

We also note that the approach of quantity over quality of documentation is preferred by many organizations. This course of action will lead to a general disillusionment with the validation process and should be avoided at all costs. A good installation qualification should fit on one or two sides of a page (three with boilerplate, if that is unavoidable). A user requirements specification (URS) may be no more than a paragraph. While it is plausible that a URS may turn out to be several hundred pages for an internally developed repository system, that would be an exception rather than the rule. If an FDA inspector arrives with a method to determine the mass of your documentation, rather than with a desire to view the processes that such paperwork documents, it will be time to find other sources of advice on validation.

3.4.1 User Requirements Specification (URS)

The key document to be generated solely by the pharmacometrician is the user requirements specification. The URS simply states the purpose of the software. It is quite worthwhile to note what the URS is *not*. For example, the business process (or scientific process) that is being addressed by the software should have already been addressed in the SOPs relevant to the department and should not appear in the URS. That is, describing what you do and the generic manner in which you do it is the fodder for a good set of SOPs, not the requirements that outline a particular tool that you wish to use. Furthermore, the system requirements—software/hardware availability, user access, recovery of data—are *not* elements of the user requirements. In many cases these elements should be covered by SOPs of the IT group or in a separate systems requirements document. The document containing these requirements (see Section 3.4.2) is a document generated by the IT engineer in collaboration with the pharmacometrician.

What the URS should contain are the features and functionality of the software tool that are required by the pharmacometrician to accomplish the business/scientific objectives at hand. As an example, Table 3.3 shows some generic user specifications that might be included in the URS for a statistical package.

Note that the URS is generic; it could fit SAS as well as S-Plus or GraphPad. One need not list all the features and functionality of the package being implemented, but the key features that one will use (and therefore test) must be included.

TABLE 3.3   Possible Elements of a User Requirements Specification

Requirement	Description
Data input formats	Must be able to import .xls, .txt, SAS transport files, etc.
Data output formats	Must be able to export to .xls, .txt, SAS transport files, etc.
Other data I/O	ODBC or JDBC connectivity
Data manipulation	Ability to subset, merge, transpose, or filter (using multiple criteria) data
Reporting	Ability to integrate output into word processing software
Statistics	Descriptive, hypothesis testing, multivariate, nonparametric, etc.
Graphics	Charts, plots, and user designed graphs
Automation or customization	Standard or vendor-designed programming, macro or automation language

3.4.2   System Specification

The generation of this document requires the interaction of the pharmacometrician with the IT group. This document reflects additions to practices already defined in IT and clinical SOPs. That is, IT should already have (see Section 3.3) SOPs that set forth:

1. Data/system backup procedures for validated systems.
2. User access (Logical Security) for validated systems access.
3. Physical access (Security) for validated systems.
4. Disaster recovery plans for validated systems.
5. Installation requirements for hardware and operating systems used for validated systems.

The main purpose of this document is to reflect the input of the IT professional regarding the system requirements, usually as documented by the vendor of the software. This clearly involves assuring that the information technologist has become familiar with the vendor's installation procedure and requirements. The selection or identification of hardware cannot proceed until the IT professional has ensured that the appropriate processor, disk space, and communication interfaces exist as required by the vendor. Furthermore, it is plausible that the software system will need software interfaces (such as database connectivity), which require additional resources. This document might reflect the "coupling" of validated systems (i.e., obtaining clinical data from a clinical database for PK/PD analysis). The document also might reflect the use of a hardware system (i.e., a user workstation) already validated for use as the target system of the new application.

The primary responsibility for generation of this document lies with the IT group. Since these are the experts at systems implementation, it behooves the pharmacometrician to engage these resources early and to try as best as possible to understand the constraints, both technical and political, under which these colleagues may be operating.

3.4.3 Validation Plan

As discussed earlier, the validation plan is a document that should be well defined by existing SOPs written within either the Quality Assurance Organization or the IT group. Unless the project is quite unique, the validation plan should follow the same general course. The user requirements and system specifications documents will impact this, of course, but the generation of this document, which also should be quite brief, should be straightforward. The key elements of the validation plan are given in Table 3.4.

Once again, sections of the validation plan regarding security, access, and so on may be better covered in SOPs that are resident with the IT group, rather than being specified for each validation.

3.4.4 Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ)

One analogy used to describe the function of these three processes is the installation, operation, and performance of an overhead projector. The IQ involves receiving the projector from the vendor, unpacking it according to the vendor's instructions, setting it on a cart or table (consistent with the vendor's requirements regarding how strong a table or cart), putting together the projector arm and head, plugging the projector into an electric socket, and turning on the power. Assuming the projector comes on, following the vendor's recommended shut-down procedure (i.e., making sure the cooling fan stays on for some fixed time after the bulb has been turned off) successfully would imply a successful IQ.

The OQ would then involve turning on the projector, taking a standard, widely used transparency, placing it on the glass, adjusting the height, distance, and focus of the projector and projector head, and so on until a satisfactory image is obtained on an acceptable image surface (i.e., a screen). Finally, the PQ would require the same type of process as in the OQ and that could be successfully performed *on the end user's* specific viewgraphs, be they color, black and white, multiple levels, or partially blocked.

At any stage in these processes, there needs to be an ordered set of steps and tests that verify the successful execution of the intended actions. This is referred to as the test script (or test plan). For each document that describes one of these qualification processes, the test script is the main functional part of the document. The "boilerplate," describing the project, referencing the validation plan, and documenting who is executing the qualification, could be as small as a single page (or even paragraph).

TABLE 3.4 Key Elements of a Validation Plan

Overview of the system
Definition of the system: user requirements, system requirements, and software description
Organization and responsibilities of the validation team (usually the end user, and the information technology and quality assurance members)
Outline of timeframe for implementation
Documentation: URS, SRS, IQ, OQ, PQ, change control, acceptance

The IQ test script clearly needs to be generated by the individual who is responsible for the installation. Usually this will be an information technologist. Note that this individual must be familiar with the software (i.e., the individual has read the installation instructions and warnings provided by the vendor) and will have to document the steps that will be followed. It is highly recommended that while this area is not the specific responsibility of the end user, the pharmacometrician would be wise to become familiar with the installation process. Vendors often provide a good deal of related information that the information technologist either misses or does not understand the analytical implications of, and it is best if the end user asks as many questions as possible before the IQ is generated.

The OQ test script may be written by either the information technologist or the pharmacometrician. The ideal situation is for this to be a collaborative effort. One highly positive result of such a collaboration is that the OQ test can turn into the best "software training" experience that both individuals will have for the particular software involved. The need to actually read the vendor's user manual in order to generate meaningful test scripts can lead to an unanticipated benefit of identifying software capabilities that were previously unknown.

There is one school of thought that claims that all of the features, functionality, buttons, menus, and so on of a particular software package must be exercised in order to successfully test the operation of the software. In general, this is extreme. Almost all of the software that is purchased has been quality assured by the vendor. Assuming that the software vendor has been audited (or that the customary use of the software by industry and regulatory agencies is widespread and it is generally agreed that the software is of high quality) and there is documented vendor evidence of functional testing, the OQ can generally be executed based on recommended tests provided by the vendor, in addition to statistical testing provided by standards organizations (15). In Section 3.5 some specific examples are outlined.

The OQ also needs to test some of the system specification requirements. These include security (i.e., authorized users can access the software, unauthorized users cannot), recovery (the software can be reinstalled and critical data recovered from original media or backup systems in the event of either accidental or disaster-related events), and boundary tests (e.g., maximum users allowed, maximum data set size).

Finally, the PQ will execute some of the same tests performed in the OQ, but using the particular functionality (noncompartmental and/or compartmental models, statistical tests, graphics, integrations, fitting) of the software that is particular to the uses of the pharmacometrician. These tests should be performed on actual data or at least data that is indicative of that analyzed during the PK/PD analysis. As in any well designed scientific investigation, this will involve the use of estimation (perhaps using other tools), boundary testing, and calibration with data standards in order that the pharmacometrician is confident that the result is "reasonable." Clearly, this is the domain of the scientist. The IT and quality assurance resources may be available to help with execution of a performance qualification, but ultimately the design responsibility for these tests lies with the pharmacometrician. If a particular type of analysis is common (i.e., bioavailability–bioequivalence–drug interaction), often the vendor or provider (i.e., for software originating in academic venues) of the software has a canonical example for the particular type of analysis. This may

be used as a model for the PQ testing, with both the vendor's and the end user's data being used to validate the algorithmic approach and result.

In the case of new types of analysis that are developed after the software has been qualified, it is incumbent upon the scientist to follow a similar process of estimation or validation in order to document the validity of the approach. This can be documented in a separate SOP (particularly if the approach becomes widely used within the organization) rather than "requalifying" the software. A change control document can be issued to indicate to the quality assurance group that "new" functionality is being employed within the software package. During the next "requalification" or upgrade of the software package, the new analytical approach can be integrated into a revised PQ.

3.5 INFORMATIVE EXAMPLES

The outlines of typical test scripts for an IQ, OQ, and PQ appear in Appendixes 3.1–3.3. Note that the "boilerplate" for these documents will be determined to a great extent by the quality assurance group. The main points to note are that each script provides a general outline of what will be tested, a statement as to responsible parties, and then a sequence of test steps that must be followed, verified, and documented as to anomalies or unexpected results. In some steps figures are called for. These are location specific and have not been reproduced here. If there are unexplained events that cannot be corrected and documented during the test, it may be necessary to regenerate the test script (maintaining the original test data as an appendix to the validation documents) and retest. We now discuss useful starting points for operational qualification scripts for various PK/PD analysis tools.

ADAPT II   There are several sample tests provided by D'Argenio and Schumitzky (16). The Fortran compiler is a key software subsystem for both ADAPT II and NONMEM. In this regard it is best to have a separate qualification for the installation of the compiler, followed by careful review of the expected output provided in Ref. 16. Older Fortran f77 compilers may show discrepancies that can only be resolved by implementing the most current versions of the f77 compiler.

NONMEM   For the operational qualification, a careful review of the parameters discussed in Section 2.9 of the *NONMEM Users Guide—Part III* (17) should be performed. These values should be identified and set during the IQ and tested properly during the OQ. The specific examples provided for NONMEM's PREDPP, NM-TRAN, and associated library subroutines are highly recommended as a starting point for the OQ. The Phenobarbital and Theophylline data files provided with the software (18) offer even more extensive testing appropriate (with modification) for a PQ. The output is well documented and individuals may seek to modify or parameterize the examples for their needs.

S-Plus   The `validate( )` function (19) is particularly appropriate for use during the OQ. As with NONMEM, the system settings and systemwide user parameter files (20) should be identified and implemented during the IQ. As with any statistical package, it is highly recommended that appropriate statistical analyses from standards organizations (15) be utilized as appropriate to the organization.

SAS The SAS Institute support organization has recently published resources for both validation (21) and actual IQ/OQ guidance (22). This should certainly be reviewed as a plausible starting point for the OQ. The same advice regarding analyses from standards organizations (15) applies. Please note that both S-Plus and SAS provide a wide range of capabilities for model creation, data analysis, presentation, and interfacing to databases and other software. It is incumbent on the user community to identify, at least initially, the capabilities that will be utilized in the user requirements documentation. Such software-specific capability should then be appropriately tested in the OQ.

WinNonlin WinNonlin comes with a well documented set of exercises (23) that can be used as the basis of an OQ. These exercises, as well as several additional tests, can also be obtained as an automated test package (24). This is quite useful if several installations of the product are being validated on independent workstations, or if it is anticipated that frequent requalification (due to product updates or releases) will be needed. There is a significant initial investment of time that must be made in order to learn and utilize the automated package. There may also be issues surrounding whether automated test software in itself must be qualified. Nevertheless, for those organizations willing to invest the effort, such testing is without a doubt more rigorous (and quite rapid) once implemented. As with other tools, WinNonlin provides the capability to create new model strategies with user-generated code as well as the ability to highly automate software functionality (25). As with other tools, the ability to write software for new modeling strategies adds the requirement that a SDLC process be in place for the pharmacometrician to adhere to.

Other PK/PD Software Kinetica, WinNonMix, and Trial Simulator® are examples of other software tools that may be utilized within PK/PD organizations. Each of these products provide example tests (26–28) that may form the basis of the OQ. In many circumstances, it will be difficult to anticipate the full range of use of some tools. Nevertheless, the vendor documentation generally provides a wide range of examples of functionality, which can be incorporated into an OQ.

Repository Systems Several software systems (29–31) (PKS®, EP2®, SAS Drug Development) have been released in the last several years, which enable the pharmacometrician to store PK/PD data, analyze such data in several ways, and then perform various reporting tasks (including data/results mining) across a wide variety of projects, studies, and so on. While each of these products have virtues and weaknesses, the fundamental issue that must be addressed by the clinical pharmacology community prior to considering the use of a specific system is: *How does this software fit our current processes?* Many times the need to answer this question leads to a major effort to define just what the current processes are! These systems require a high degree of organizational discipline and structure around the concept of metadata. That is, what *data do we use to describe* the models, data, analysis results, and reporting variables that are critical to our organization?

The important point to recognize here is that software systems such as "repository" systems are considered "enterprise" software. The implementation is not customizable to an individual's requirements or a department's needs. The architecture of the software is the vendor's "impression" of how a clinical pharmacology effort

may be organized. This "impression" may have no connection to your current processes; it may, especially if one's organization was the model used when the software was architected, reflect your processes exactly.

Under any circumstances, the implementation of such a package requires a large effort to identify processes, especially processes between groups such as data management, biostatistics, quality assurance, and clinical pharmacology, before considering individual software systems. Recent analyses of enterprise software have characterized this effort as the "organizational capital" (32) that must be expended in addition to the resources for "capital equipment and software expense." Once a system is chosen, the implementation team needs to recognize that the fundamental way they work will be changed. The rewards may be tremendous, but the road to implementation may be long and arduous.

3.6 SUMMARY

This chapter is a brief attempt to aid the pharmacometrician in understanding how "quality" standards need to be applied to research and development activities involving software tools. Specifically, the needs of the ethical pharmaceutical industry are addressed, but one could argue that the ability to document such activities is critical in any industry. In the chapters that follow, several specific analytical approaches to numerous problems in clinical pharmacology are discussed. If the software tools utilized in these creative and important analytical methodologies are properly installed, validated, and supported, the quality and throughput of the pharmaceutical realization process will be assured, for both the development teams and the regulatory agencies involved in the process of ethical drug discovery and development.

REFERENCES

1. T. Stokes, R. C. Branning, K. G. Chapman, H. J. Hambloch, and A. J. Trill, *Good Computer Validation Practices: Common Sense Implementation.* Interpharm Press, Buffalo Grove, IL, 1994.

2. R. Chamberlain, *Computer Systems Validation for the Pharmaceutical and Medical Device Industries.* Alaren Press, Libertyville, IL, 1994.

3. FDA, `http://www.fda.gov/ora/compliance_ref/part11/`.

4. *FDA: Help Us Plot Act Two for Part 11.* `http://www.bio-itworld.com/news/051904_report5155.html`.

5. ISO, `http://www.iso.org/iso/en/iso9000-14000/iso9000/iso9000index.html`.

6. ANSI, Document ANSI/IEEE 730-2002, *Software Quality Assurance Plans*; also see `www.ansi.org`.

7. IEEE, Document 1061–1998, *IEEE Standard for a Software Quality Metrics Methodology*; also see `http://standards.ieee.org/catalog/olis/se.html`.

8. G. J. Myers, *Software Reliability: Principles and Practices.* Wiley, Hoboken, NJ, 1976.

9. US EPA, *Good Automated Laboratory Practices: Recommendations for Ensuring Data Integrity in Automated Laboratory Operations with Implementation Guidance.* US Environmental Protection Agency, Research Triangle Park, NC, 1990.

10. FDA, *21CFR Part 312: New Drug Product Regulations; Final Rule.* Department of Health and Human Services, Washington, DC; *Fed Reg* 3/19/1987, Part VII.

11. FDA, *Guide to Inspection of Computerized Systems in Drug Processing.* US Government Printing Office, Washington, DC, 1983-381-166:2001, 1983.

12. FDA, *Technical Report on Software Development Activities: Reference Materials and Training Aids for Investigators*, US Government Printing Office, Washington, DC, July 1987.

13. An internet search for "21 CFR 11" or "21CFR11" using www.google.com retrieved 9550 results at the time of this writing.

14. B. D. McCullough and B. Wilson, On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Comput Statistics Data Analysis* **40**:713–721 (2002).

15. NIST Statistical Reference Datasets, www.itl.nist.gov/div898/strd.

16. D. Z. D'Argenio and Alan Schumitzky, *ADAPT II Pharmacokinetic/Pharmacodynamic Systems Analysis Software User's Guide to Release 4.* The Biomedical Simulations Resource, University of Southern California, Los Angeles, Mar. 1997, Chapters 2, 5, and 6.

17. A. J. Boeckmann, S. L. Beal, and L. B. Sheiner, *NONMEM Users Guide—Part III.* NONMEM Project Group, University of California at San Francisco, May 1999, pp. 25–30.

18. A. J. Boeckmann, S. L. Beal, and L. B. Sheiner, *NONMEM Users Guide—Part III.* NONMEM Project Group, University of California at San Francisco, May 1999, pp. 102–104.

19. *S-Plus 6 for Windows Programmer's Guide.* Insightful Corporation, Seattle, WA, July 2001, Chapter 24.

20. *S-Plus 5 for Unix Installation and Maintenance Guide.* Data Analysis Products Division, Insightful Corporation, Seattle, WA, May 1999, Chapters 4 and 7.

21. SAS Corp., http://support.sas.com/rnd/migration/planning/validation/.

22. SAS Corp, http://support.sas.com/rnd/migration/resources/SAS_IQOQ_SOP.pdf.

23. *WinNonlin® Getting Started Guide Version 4.1.* Pharsight Corporation, Mountain View, CA, 2003, Chapter 3.

24. Pharsight, http://www.pharsight.com/products/winnonlin/wnl_validation.php.

25. *WinNonlin® Users Guide Version 4.1.* Pharsight Corporation, Mountain View, CA, 2003, Chapters 11, 12, 19, and 20.

26. *Kinetica® R4.0 User Manual V4.0.* InnaPhase Corp., Philadelphia, PA, 2001, pp. 256–301.

27. *WinNonMix® Getting Started Guide Version 2.0.1.* Pharsight Corporation, Mountain View, CA, 2000, Chapter 2.

28. *Pharsight® Trial Simulator™ Getting Started Guide Version 2.1.2.* Pharsight Corporation, Mountain View, CA, 2001, Chapter 4.

29. Pharsight Corp, http://www.pharsight.com/products/pks_suite/.

30. Innaphase Corp, http://www.innaphase.com/products_epseries.html.

31. SAS Corp, http://www.sas.com/industry/pharma/develop/index.html.

32. Federal Reserve Bank of Minneapolis Research Department Staff Report 291, http://research.mpls.frb.fed.us/research/sr/sr291.pdf.

APPENDIX 3.1   SAMPLE INSTALLATION QUALIFICATION

Document Number: IQ00-00001 Document Location: QA Document Owner: Technolgist, Info	**Install Qualification:** **S-Plus 6.0.4**	Test Version: 1.0 Reference Document(s) #: SV00-00001

Purpose
(This defines the purpose for the current test(s). It may refer to previous tests and/or documents.) This test is conducted to qualify the installation of software components used for S-Plus 6.0.4. These are components of the S-Plus 6.0.4 software and are used to verify that a client interface may access S-Plus 6.0.4. These tests will also be used to establish a baseline for future testing.

Scope
(This defines the scope of the test(s). It is a written description of the scope and restrictions of the test.) Testing is done to prove the S-Plus 6.0.4 software has been correctly installed. This test does not prove the Installation Process; instead, it proves that the end result of the process was successful based on software functionality. Validating the result of the installation implicitly proves the success of the process.

Test Requirements
Testing is done to prove the following:

1. Verify the S-Plus 6.0.4 installation
    1.1. Verify the my_server_name server is started
    1.2. Verify the */home/splus6* directory and permissions
    1.3. Verify the S-Plus scripts have been copied to my_server_name: */usr/local/bin*
    1.4. Verify the file and directory listing for the */home/splus6* directory
    1.5. Verify that */usr/local/bin/Splus and /usr/local/bin/Splus* invoke S-Plus 6.0.4
2. Verify Terminal/HOST Client/Server Interface
    2.1. Verify UNIX Server login from Telnet client (terminal)
        2.1.1. Verify Security
            2.1.1.1. Bad Username
            2.1.1.2. Good Username, Bad Password
    2.2. Verify Logout from Telnet client (terminal)
    2.3. Verify UNIX Server login from X-windows client
        2.3.1. Verify Security
            2.3.1.1. Bad Username
            2.3.1.2. Good Username, Bad Password
    2.4. Verify Logout from X-windows client

Test Prerequisites
(A list of requirements necessary to run the test. These can include environment reqs (e.g., NT, with MS Office loaded), tester reqs (e.g., tester is trained in operating MS Office), software reqs (e.g., test assumes xyz software to have already been loaded), or other reqs (e.g., paper documents for scanning).) The following conditions must be met before testing:

- The environment is ready to test.
- Tester is trained in basic usage of UNIX and S-Plus 6.0.4.
- Testing files are prepared and put in place.

Test Instructions

(Gives any special instructions to the tester. Tester is assumed to be qualified to execute test.) For each test condition in the Testing Table, the Tester must initial each graybar section when completed regardless of success or nonsuccess. If the test condition has been met and Expected Result is the same as the actual result (the result of executing the test condition), then the test is successful and must be marked as OK in the OK column. If the test condition has not been met, or the Expected Results are not exactly the same as the actual results, then the Tester must stop, report the deviation in the Comments column, and report the occurrence to the Test Coordinator. At that time, the Test Coordinator will make a judgment on whether or not the test can be continued. In the event that the deviation is considered acceptable and that the test can continue, the Test Coordinator must log the event, any workarounds necessary, and initial the Tester's comments (this may be done on the script if there is room). In the event that the deviation is not acceptable, then the test must stop.

Test Tables

Test tables show:

1. Line Number: Allows reference for tracking anomalies and errors.
2. Test Condition: Defines the test.
3. Expected Results: Defines what should happen. Any deviation is an error.
4. OK: Were the expected results met?
5. Initials: Tester proof of execution.
6. Comments: Allows information about the test condition.

Signoffs

(Signoffs for Document Owner and Author with printed name and date spaces.)

Author **(By signing this the author of this document agrees that the document is complete and accurate.)**

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Author	

Owner **(By signing this the owner of this document agrees that the document is complete and accurate.)**

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Owner	

#	Test Condition	Expected Results	OK	Init	Comments
1. Verify the S-Plus 6.0.4 Installation					
1.1. Verify the my_server_name server is started and accessible					
1.	As a valid my_server_name user and member of the clinical group, open a Telnet client and log on to my_server_name server.	Login prompt and shell prompt observed following successful login as in Figure 1.			
2.	Verify the home directory by typing *pwd.*	System responds with user's home directory, */home/"username"*			
1.2. Verify the */home/splus6* directory and permissions					
3.	Type *cd /home* Type *ls –al	grep splus*	The output is: *drwxr-x— 12 splus clinical 1024 Mar 23 11:06 splus* Indicating the *splus* directory exists and is accessible (r-x) by the group *clinical*			
1.3. Verify the S-Plus scripts have been copied to my_server_name: */usr/local/bin*					
4.	Type *cd /usr/local/bin* Type *ls –l*	The output is: -rwxr-x— 1 root clinical 4655 Sep 5 2004 Splus -rwxr-x— 1 root clinical 4655 Mar 22 14:32 Splus			
1.4. Verify the file and directory listing for the */home/splus6* directory					
5.	Type *cd /home/splus6* to return to the */home/splus6* directory. Type *ls –lR	pg*	Compare the output of this command to Figure 2. ALL of the files listed in Figure 2 *must* be in the output. There will be files output by the command *that are not* displayed in Figure 2.			

72

1.5. Verify that */usr/local/bin/Splus* and */usr/local/bin/Splus* invoke S-Plus 6.0.4			
6.	Type *cd $HOME* Type */usr/local/bin/Splus*	The output file should be similar to Figure 3. There should be no indication of errors. The line: "Working data will be in /home/wolk/MySwork" should indicate the current "username" instead of "wolk."	
7.	Type *q()* at the > prompt. Type */usr/local/bin/Splus*	The operating system command prompt should return after the first command. The second command should produce the same result as Step 6. Type *q()* to exit S-Plus.	
2. Verify Terminal/HOST Client/Server Interface			
2.1. Verify UNIX Server login from Telnet client (terminal)			
2.1.1. Verify Security			
2.1.1.1. Bad username			
8.	As a valid my_server_name user and member of the pharmaco group, open a Telnet client and log on to my_server_name server using a bad user name, good password.	Login is rejected as in Figure 4.	
2.1.1.2. Good username, bad password			
9.	As a valid my_server_name user and member of the pharmaco group, open a Telnet client and log on to my_server_name server using a good user name and an invalid password.	Login is rejected as in Figure 5.	

2.2. Verify Logout		
10.	As a valid my_server_name user and member of the pharmaco group, open a Telnet client and log on to my_server_name server using a good user name and a valid password. Then type exit.	Login is successful as in Figure 1. Telnet client screen is as in Figure 6 and keyboard is unresponsive following exit. Logout is successful.
2.3. Verify UNIX Server login from X-windows client		
2.3.1. Verify Security		
2.3.1.1. Bad username		
11.	Configure X-windows client to connect to my_server_name server (using eXodusPowerPC, select "Connections/ Connection Manager/Sample XDM Session/Edit" Change "Mode" to "Query" set "Host" to "my_server_name" and "Title" as "my_server_name IQ"; Click connect using an invalidusername.	Login is not allowed as in Figure 7. If using the Common Desktop Environment instead of Openwindows, the windows may appear differently than in the figures below, but the functionality should be the same.
2.3.1.2. Good username, bad password		
12.	Login to configured X-windows client using a valid username / invalid password.	Login is not allowed as in Figure 8.
2.4. Verify Logout		
13.	Login to configured X-windows client as a valid my_server_name user and member of the pharmaco group, log on to my_server_name server using a good user name and a valid password. Then select Mouse Button 3; Click on Exit.	Valid login produces X-windows, Openwindows environment similar to Figure 9. Selecting Mouse button 3 yields window similar to that in Figure 10. Selecting exit yields confirmation screen similar to Figure 11. Clicking on Exit causes windows environment to disappear.

Signoffs

(Signoffs for Document Tester and Test Coordinator with printed name and date spaces)

Tester   (By signing this the tester of this document agrees that the test has been completed and is accurate.)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Tester	

Test Coordinator   (By signing this the tester of this document agrees that the test has been completed and is accurate.)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Test Coordinator	

APPENDIX 3.2   SAMPLE OPERATION QUALIFICATION

Document Number: OQ00–00001 Document Location: IS Validation Document Owner: Metrician, Pharmaco	**Operation Qualification: S-Plus 5.1** **Appendix A**	Test Version: 1.0 Reference Document(s) #: SV00- 00001

Purpose

(This defines the purpose for the current test(s). It may refer to previous tests and/or documents.) This test is conducted to qualify the operation of S-Plus 5.1. These tests verify the proper operation of the S-Plus 5.1 software as well as the security of the software as far as user access to the data and executable programs. It also tests the backup and recovery of data and executables. These tests will also be used to establish a baseline for future operational testing.

Scope

(This defines the scope of the test(s). It is a written description of the scope and restrictions of the test.) Testing is done to prove the S-Plus 5.1 is operational in accordance with the manufacturer's criteria. Testing is also done to prove that data and executable software access is limited to authorized users and only up to the number of available licenses. Testing is also done to verify the backup and restoration of selected data and executable files.

Test Requirements

Testing is done to prove the following:

1. Verify the S-Plus 5.1 Operation (these tests are from the S-Plus 2000 Programmer's Guide, Chapter 25)
   1.1. Execute *validate()*; the complete validation test suite
   1.2. Verify the *validate* function code
   1.3. Execute the *anova* test suite in verbose mode to demonstrate an individual test
   1.4. Execute the *regress* test suite in verbose mode, return and examine the Boolean result
2. Verify the S-Plus Data Files May only be accessed by Authorized Users
   2.1. Verify /home/splus directory is not accessible to unauthorized users
   2.2. Verify the /home/splus and /home/"user" subdirectories are (not) writeable by (unauthorized) authorized users
   2.3. Verify the /home/splus and /home/"user" subdirectories are readable by the authorized group
3. Verify the S-Plus program may be started only by authorized users
   3.1. Verify S-Plus or S-Plus 5 may be started only by authorized users
   3.2. Verify that the license limit may not be exceeded by authorized users.
4. Verify that a tape backup of data and executable files may be selectively restored

Test Prerequisites

(A list of requirements necessary to run the test. These can include environment reqs (e.g., NT, with MS Office loaded), tester reqs (e.g., tester is trained in operating MS Office), software reqs (e.g., test assumes xyz software to have already been loaded), or other reqs (e.g., paper documents for scanning).) The following conditions must be met before testing:

- The environment is ready to test.
- Tester is trained in basic usage of UNIX and S-Plus 5.1.
- Testing files are prepared and put in place.

Test Instructions

(Gives any special instructions to the tester. Tester is assumed to be qualified to execute test.) For each test condition in the Testing Table, the Tester must initial each graybar section when completed regardless of success or nonsuccess. If the test condition has been met and Expected Result is the same as the actual result (the result of executing the test condition), then the test is successful and must be marked as OK in the OK column. If the test condition has not been met, or the Expected Results are not exactly the same as the actual results, then the Tester must stop, report the deviation in the Comments column, and report the occurrence to the Test Coordinator. At that time, the Test Coordinator will make a judgment on

whether or not the test can be continued. In the event that the deviation is considered acceptable and that the test can continue, the Test Coordinator must log the event, any workarounds necessary, and initial the Tester's comments (this may be done on the script if there is room). In the event that the deviation is not acceptable, then the test must stop.

Test Tables
Test tables show:

1. Line Number: Allows reference for tracking anomalies and errors.
2. Test Condition: Defines the test.
3. Expected Results: Defines what should happen. Any deviation is an error.
4. OK: Were the expected results met?
5. Initials: Tester proof of execution.
6. Comments: Allows information about the test condition.

Signoffs
(Signoffs for Document Owner and Author with printed name and date spaces.)

Author   (By signing this the author of this document agrees that the document is complete and accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Author	

Owner   (By signing this the owner of this document agrees that the document is complete and accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Owner	

#	Test Condition	Expected Results	OK	Init	Comments
1. Verify the S-Plus 5.1 Operation (these tests are from the S-Plus 2000 Programmer's Guide, Chapter 25)					
1.1. Execute *validate()*; the complete validation test suite					
1.	As a valid S-Plus user (Section 7.2 of Operation Qualification) open a Telnet or X-windows client and log on to my_server_name server. Verify the current directory is /home/"username" by typing *pwd*.	Successful login to my_server_name server *pwd* command returns: /home/"username" (for ksh or sh) /export/home/"username" (for csh) where username is the tester's login name			
2.	Type: */usr/local/bin/Splus*	System responds with: License Warning : S-Plus license expires Thu Apr 20 23:59:59 2009 S-Plus: Copyright © 1988, 1999 Insightful, Inc. S : Copyright Lucent Technologies, Inc. Version 5.1 Release 1 for Sun SPARC, SunOS 8 : 2009 Working data will be in /home/myname/MySwork > Except username is user of Step 1. Instead of "myname"			
3.	At the > prompt, type: *validate()*	System responds (the test may take a few minutes) with the results depicted in Figure 1.			
1.2. Verify the *validate* function code					
4.	At the > prompt, type: *sink("/home/***username***/validate_code")* where **username** is the user name observed in the *pwd* command of Step 1. At the > prompt, type: *validate*	System responds with output of Figure 2.			

#	Action	System response	
	Open a second Telnet session as in Step 1 and Login to that session. In the second session type: *pg validate_code*		
1.3.	Execute the *anova* test suite in verbose mode to demonstrate an individual test		
5.	In the first Telnet session, type the following: *sink("/home/***username*** /annova_test")* where **username** is the user name observed in the *pwd* command of Step 1. At the > prompt, type: *validate("anova", verbose = T)* In the second Telnet session type: *pg anova_test*	System responds with output of Figure 3.	
1.4.	Execute the *regress* test suite in verbose mode, return and examine the Boolean result		
6.	In the first Telnet session, type the following: *sink("/home/***username*** /regress_test")* where **username** is the user name observed in the *pwd* command of Step 1. At the > prompt, type: *Regrspass < -validate("regress", verbose = T)* In the second Telnet session type: *pg regress_test*	System responds with output of Figure 4.	
7.	At the > prompt, type: *sink()* *Regrspass*	System responds with the Boolean result "T" for True: [1] T >	

	2. Verify the S-Plus Data Files May only be accessed by Authorized Users		
	2.1. Verify the /home/splus directory is not accessible to unauthorized users		
8.	As a valid my_server_name user who is NOT listed as a S-Plus user (Section 7.2 ofOperation Qualification), open a Telnet session and log on to my_server_ name (The password for the "nmtest" user may be obtained from the test coordinator). Verify the current directory is /home/"username" by typing *pwd*.	Successful login to my_server_name *pwd* command returns: /home/"username" where username is the tester's login name	
9.	Type: *cd /home/splus*	Access is denied, message is: ksh: /home/splus: permission denied or for csh: /home/splus: Permission denied	
	2.2. Verify the /home/splus and /home/"user" subdirectories are (not) writeable by (unauthorized) authorized users		
10.	Type: *mkdir /home/splus/test_dir*	System responds with: mkdir: Failed to make directory "/home/splus/test_dir"; Permission denied	
11.	Type: *mkdir /home/wolk/test_dir*	System responds with: mkdir: Failed to make directory "/home/wolk/test_dir"; Permission denied	
12.	Type: *exit*	Successful logoff from my_server_ name server	
13.	As a valid S-Plus user (Section 7.2 of Operation Qualification), open a Telnet session and log on to my_server_name.	Successful login to my_server_name	

14.	Type: *pwd*	*pwd* command returns: /home/"username"		
15.	Type: *mkdir /home/splus/test_dir*	System responds with: mkdir: Failed to make directory "/home/splus/test_dir"; Permission denied NOTE THAT AUTHORIZED USERS HAVE READ AND EXECUTE BUT NOT WRITE PERMISSION in /home/splus		
16.	Type: *mkdir $HOME/test_dir* *ls –l $HOME	grep test_dir*	A new directory named test_dir has been written by the current user to their HOME directory, as demonstrated by the response of the *ls –l* command: drwxr-xr-x 2 wolk pharmaco 512 Mar 20 09:13 test_dir (Note that the user who logged on in step 13 should be listed instead of "wolk")		
17.	Type: *rmdir $HOME/test_dir* *ls –l $HOME	grep test_dir*	The *ls –l* command returns nothing after the directory has been removed.		
18.	Type: *mkdir /home/wolk/test_dir*	System responds with: mkdir: Failed to make directory "/home/wolk/test_dir"; Permission denied (Group members directories are NOT writeable by other group members.)		

2.3	**Verify the /home/splus and /home/"user" subdirectories are readable by authorized group**	
19.	Type: *cd /home/splus* *cat Copyright*	File contents of the *Copyright* file appear as in Figure 5.
20.	Type: *cd /home/wolk pwd*	Output of the *pwd* command is: /export/home/wolk (csh) /home/splus (ksh or sh)
21.	Type: *cat README*	File contents of the README file appear as in Figure 6. Group members files are readable by other group members unless read permission is specifically removed.
3. Verify the S-Plus program may be started only by authorized users		
3.1.	**Verify S-Plus or S-Plus 5 may be started only by authorized users**	
22.	Type: */usr/local/bin/Splus5* The S-Plus command was tested in Step 2. That step may be repeated here if so desired.	System responds with: License Warning : S-Plus license expires Thu Apr 20 23:59:59 2000 S-Plus : Copyright (c) 1988, 1999 MathSoft, Inc. S : Copyright Lucent Technologies, Inc. Version 5.1 Release 1 for Sun SPARC, SunOS 5.5 : 1999 Working data will be in /home/wolk/MySwork > Except username is user of Step 1. Instead of "wolk."
23.	Type *exit*	Successful logoff from my_server_name.

24.	As a valid my_server_name user who is NOT listed as a S-Plus use r (Section 7.2 ofOperation Qualification), open a Telnet session and log on to my_server_name (The password for the "nmtest" user may be obtained from the test coordinator). Verify the current directory is /home/"username" by typing *pwd*.	Successful login to my_server_name *pwd* command returns: /home/"username" where username is the tester's login name.			
25.	Type: */usr/local/bin/Splus*	System responds with, for kshell: ksh: /usr/local/bin/Splus: cannot execute for c-shell: /usr/local/bin/Splus: Permission denied			
26.	Type: */usr/local/bin/Splus5*	System responds with, for kshell: ksh: /usr/local/bin/Splus: cannot execute for c-shell: /usr/local/bin/Splus: Permission denied			
27.	Type *exit*	Successful logoff from my_server_name.			
3.2 Verify that the license limit may not be exceeded by authorized users.					
28.	Repeat Steps 1 and 2 above. For the remainder of this test, a second valid user of S-Plus must be utilized. CONTACT THE TEST COORDINATOR IN ORDER TO CONTINUE THIS TEST! Have the test coordinator repeat Steps 1 and 2 above.	System output should resemble that in Figure 7. The "username" should be the tester's username. The test coordinator is not allowed to execute S-Plus due to the license limit: *Terminating S Session: No S-Plus licenses available*			
29.	Type *exit*	Successful logoff from my_server_name.			

4. Verify that a tape backup of data and executable files may be selectively restored				
30.	Login to my_server_name as kroot (this testmust be executed by a System Administrator or an Application Administrator with System Administrator Privilege).	Login successful.		
31.	Rename the file /usr/local/bin/Splus5 to Splus5_orig; Repeat Step 22.	An error message occurs, the executable shell script is not found.		
32.	Obtain a backup tape of the my_server_name server. Restore the file/usr/local/bin/Splus5 to it's original location. Repeat Step 22.	System response is: License Warning : S-Plus license expires Thu Apr 20 23:59:59 2000 S-Plus : Copyright © 1988, 1999 MathSoft, Inc. S : Copyright Lucent Technologies, Inc. Version 5.1 Release 1 for Sun SPARC, SunOS 5.5 : 1999 Working data will be in /home/wolk/MySwork > Except username is "kroot" instead of "wolk"		

33.	Delete the restored version; rename the original version to S-Plus 5. Verify that the directory /home/kroot exists and that the file /home/kroot/ MySwork exists within this directory. Type: *ls –l /home/kroot	grep MySwork*	Output of *ls –l /home/kroot	grep MySwork:* drwxr-xr-x 3 root other 512 Sep 5 1999 MySwork		
34.	Type: *rm -r /home/kroot/MySwork ls –l /home/kroot/MySwork	grep MySwork*	Output of *ls –l /home/kroot	grep MySwork* Is empty.		
35.	Repeat Step 22.	S-Plus generates the message in Figure 8, which, because the MySwork directory was deleted in Step 34, includes: *Creating data directory for chapter /home/kroot/MySwork*		
36.	Repeat Step 34. Obtain a backup tape of the my_server_name /home/kroot directory. Restore the directory /home/kroot/MySwork to it's original location. Repeat Step 22.	S-Plus generates the message in Figure 9, which since the MySwork directory was restored, does not have the "Creating data. . ." message of Figure 8.		
37.	Type *exit*.	Successful logoff from my_server_name.		

Signoffs

(Signoffs for Document Tester and Test Coordinator with printed name and date spaces:)

Tester   (By signing this the tester of this document agrees that the test has been completed and is accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Tester	

Tester   (By signing this the tester of this document agrees that the test has been completed and is accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Tester	

Test Coordinator   (By signing this the tester of this document agrees that the test has been completed and is accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Test Coordinator	

APPENDIX 3.3   SAMPLE PERFORMANCE QUALIFICATION

Document Number: PQ00–00001
Document Location: Quality
   Assurance
Document Owner: Metrician,
   Pharmaco
	Performance Qualification:
	S-Plus 5.1
	
	Appendix A
		Test Version: 1.0
		Reference Document(s)
		   #: SV00-00001

Purpose

(This defines the purpose for the current test(s). It may refer to previous tests and/or documents.) This test is conducted to qualify the performance of S-Plus 5.1. These tests verify the proper operation of specific, commonly used features of the S-Plus system software.

Scope

(This defines the scope of the test(s). It is a written description of the scope and restrictions of the test.) Testing is done to prove that specific, commonly used features of S-Plus 5.1 are operational in accordance with the end users needs. The data sets are obtained from the National Institute of Standards and Technology (Ref. 6) and academic reference texts (Ref. 7)

Test Requirements

Testing is done to prove the following:

1. Verify that S-Plus 5.1 performs the NIST StRD Analysis of Variance calculations to within 3 signficant digits
   1.1. Test ANOVA with dataset SiRstv, that is, low degree of stiffness, low replicates per cell
   1.2. Test ANOVA with dataset AtmWtAg, that is, average degree of stiffness, low replicates per cell
   1.3. Test ANOVA with dataset SmLs06, that is, average degree of stiffness, high replicates per cell
2. Verify that S-Plus 5.1 performs the NIST StRD Linear Regression calculations to within 3 signficant digits
   2.1. Test Linear Regression with dataset Norris, Low difficulty linear
   2.2. Test Linear Regression with dataset NoInt1, Average difficulty linear
   2.3. Test Linear Regression with dataset Filip, High difficulty polynomial
3. Verify that S-Plus 5.1 performs the NIST StRD Non-linear Regression calculations to within 3 signficant digits
   3.1. Test Nonlinear Regression with dataset Misra1a, Lower difficulty exponential
   3.2. Test Nonlinear Regression with dataset Kirby2, Average difficulty rational
   3.3. Test Nonlinear Regression with dataset MGH09, Higher difficulty rational
4. Verify that S-Plus 5.1 performs a General Additive Model with Gaussian error Distribution and identity link problem correctly to 3 significant digits

Test Prerequisites

(A list of requirements necessary to run the test. These can include environment reqs (e.g., NT, with MS Office loaded), tester reqs (e.g., tester is trained in operating MS Office), software reqs (e.g., test assumes xyz software to have already been loaded), or other reqs (e.g., paper documents for scanning).) The following conditions must be met before testing:

- The environment is ready to test.
- Tester is trained in basic usage of UNIX and S-Plus 5.1.
- Testing files are prepared and put in place.

Test Instructions

(Gives any special instructions to the tester. Tester is assumed to be qualified to execute test.) For each test condition in the Testing Table, the Tester must initial each graybar section when completed regardless of success or nonsuccess. If the test condition has been met and Expected Result is the same as the actual result (the result of executing the test condition), then the test is successful and must be marked as OK in the OK column. If the test condition has not been met, or the Expected Results are not exactly the same as the actual results, then the Tester must

stop, report the deviation in the Comments column, and report the occurrence to the Test Coordinator. At that time, the Test Coordinator will make a judgment on whether or not the test can be continued. In the event that the deviation is considered acceptable and that the test can continue, the Test Coordinator must log the event, any workarounds necessary, and initial the Tester's comments (this may be done on the script if there is room). In the event that the deviation is not acceptable, then the test must stop.

Test Tables
Test tables show:

1. Line Number: Allows reference for tracking anomalies and errors.
2. Test Condition: Defines the test.
3. Expected Results: Defines what should happen. Any deviation is an error.
4. OK: Were the expected results met?
5. Initials: Tester proof of execution.
6. Comments: Allows information about the test condition.

Signoffs
(Signoffs for Document Owner and Author with printed name and date spaces.)

Author **(By signing this the author of this document agrees that the document is complete and accurate)**

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Author	

Owner **(By signing this the owner of this document agrees that the document is complete and accurate)**

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Owner	

#	Test Condition	Expected Results	OK	Init	Comments
1. Verify that S-Plus 5.1 performs the NIST StRD Analysis of Variance calculations to within 3 signficant digits					
1.1. Test ANOVA with dataset SiRstv, that is, low degree of stiffness, low replicates per cell					
1.	As a valid S-Plus user (Section 7.2 of Performance Qualification) open a Telnet or X-windows client and log on to my_server_name server. Verify the current directory is /home/"username" by typing *pwd*.	Successful login to my_server_name server *pwd* command returns: /home/"username" (for ksh or sh) /export/home/"username" (for csh) where username is the tester's login name.			
2.	Create a subdirectory in the /home/"username". Type the following: *mkdir splus_pq* *cd splus_pq pwd*	System responds to *pwd* command with: /home/"username"/splus_pq			
3.	Type the following: *Splus CHAPTER*	System responds with: Creating data directory for chapter. Splus5 chapter splus_pq initialized.			
4.	Type the following: *mkdir datasests* *cp ~username/splus_pq/ datasets/* datasets* *cd datasets* *ls -og*	System responds with: total 476 -rw-r--r-- 1 683 May 26 15:04 AtmWtAg.csv -rw-r--r-- 1 3594 May 26 15:04 Filip.csv -rw-r--r-- 1 1922 May 26 15:04 Kirby2.csv -rw-r--r-- 1 130 May 26 15:04 MGH09.csv -rw-r--r-- 1 255 May 26 15:04 Misra1a.csv -rw-r--r-- 1 166 May 26 15:04 Misra1a1.csv -rw-r--r-- 1 80 May 26 15:04 NoInt1.csv -rw-r--r-- 1 388 May 26 15:04 Norris.csv -rw-r--r-- 1 292 May 26 15:04 SiRstv.csv -rw-r--r-- 1 216126 May 26 15:04 SmLs06.csv -rw-r--r-- 1 182 May 26 15:04 Wampler1.csv			

5.	Type the following: *cd ..* *Splus*	System responds with: S-Plus : Copyright © 1988, 1999 MathSoft, Inc. S : Copyright Lucent Technologies, Inc. Version 5.1 Release 1 for Sun SPARC, SunOS 5.5: 1999 Working data will be in Data
6.	Type: *SiRstv < - importData(file =* *"datasets/SiRstv.csv") instr1 <* *-factor(SiRstv[,"Instrument"])* *resist1 < -SiRstv[, "Resistance"]* *SiRstv.anova.1 <- aov(resist1~* *instr1)summary (SiRstv.anova.1)*	System responds with: Df Sum of Sq Mean Sq F Value Pr(F) instr1 4 0.0511463 0.01278657 1.180462 0.3494475 Residuals 20 0.2166366 0.01083183

7.	Type the following:
	SiRstv.anova.1
		System responds with:
		Call:
		  aov(formula = resist1 ~ instr1)
		Terms:
		          instr1 Residuals
		Sum of Squares 0.0511463 0.2166366
		Deg. of Freedom 4 20
		
		Residual standard error: 0.10440761
		Estimated effects are balanced

8.	Review the information in Figure 1.0a.	Figure 1.0a is the general background information for the NIST StRD Anova test. Initial the INIT column if read and understood. Figure 1.0b is a summary of the data sets referred to in Figure 1.0a.
9.	Review the information in Figure 1.1a.	Figure 1.1a is the specific model information relevant to the SiRstv data set. Initial the INIT column if read and understood. Figure 1.1b is the data set information. It is provided for reference.

No.	Instruction	Response / Notes
10.	Review the results of Step 6 and Step 7 against the certified results of Figure 1.1c.	Results match within 3 significant figures.
11.	Review the data set in the of Figure 1.1d against the data imported in Step 6. (Type SiRstv at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 1.1d.
1.2. Test ANOVA with dataset AtmWtAg, that is, average degree of stiffness, low replicates per cell		
12.	Type: *AtmWtAg < - importData(file =* *"datasets/AtmWtAg.csv") instr2* *< -factor(AtmWtAg [,"Instrument"])* *AgWt < - AtmWtAg [,"AgWt"]* *AtmWtAg.anova.1 < - aov* *(AgWt~instr2)* *summary(AtmWtAg.anova.1)*	System responds with: $\quad$ Df $\quad$ Sum of Sq $\quad$ Mean Sq $\quad$ F Value $\quad$ Pr(F) instr2 $\quad$ 1 $\quad$ 3.638340e-09 $\quad$ 3.638342e-09 $\quad$ 15.94673 $\quad$ 0.0002326844 Residuals $\quad$ 46 $\quad$ 1.049517e-08 $\quad$ 2.281560e-10
13.	Type the following: *AtmWtAg.anova.1*	System responds with: Call: $\quad$ aov(formula = AgWt ~ instr2) Terms: $\qquad$ instr2 $\qquad$ Residuals Sum of Squares $\quad$ 3.638340e-09 $\quad$ 1.049517e-08 Deg. of Freedom $\quad$ 1 $\qquad$ 46 Residual standard error: 1.510483e-05 Estimated effects are balanced
14.	Review the information in Figure 1.2a.	Figure 1.2a is the specific model information relevant to the AtmWtAg data set. Initial the INIT column if read and understood. Figure 1.2b is the data set information. It is provided for reference.

#	Instruction	Response		
15.	Review the results of Step 12 and Step 13 against the certified results of Figure 1.2c.	Results match within 3 significant figures.		
16.	Review the data set in the problem summary of Figure 1.2d against the data imported in Step 12 (Type AtmWtAg at the S-Plus prompt to see the data set, if necessary, use the *options(digits = 12)* command to view more significant digits)	The data imported matches the data in Figure 1.2d.		
1.3. Test ANOVA with dataset SmLs06, that is, average degree of stiffness, high replicates per cell				
17.	Type: *SmLs06 < - importData* *(file = "datasets/SmLs06.csv")* *treat1 <-factor(SmLs06 [,* *"Treatment"])resp1 <- SmLs06* *[, "Response"]SmLs06.anova.1 <* *- aov(resp1~treat1)summary* *(SmLs06.anova.1)*	System responds with: Df Sum of Sq Mean Sq F Value Pr(F) treat1 8 160.08 20.01 2001 0 Residuals 18000 180.00 0.01		

18.	Type the following:
	SmLs06.anova.1
		System responds with:
		Call:
		   aov(formula = resp1 ~ treat1)
		Terms:
		             treat1 Residuals
		Sum of Squares 160.08 180.00
		Deg. of Freedom 8 18000
		
		Residual standard error: 0.1
		Estimated effects are balanced		

19.	Review the information in Figure 1.3a.	Figure 1.3a is the specific model information relevant to the SmLs06 data set. Initial the INIT column if read and understood. Figure 1.3b is the data set information. It is provided for reference.
20.	Review the results of Step 17 and Step 18 against the certified results of Figure 1.3c.	Results match within 3 significant figures.
21.	Review the data set in the problem summary of Figure 1.3d against the data imported in step 17 (Type SmLs06 at the S-Plus prompt to see the data set)	The data imported matches the data in Figure 1.3d.
2. Verify that S-Plus 5.1 performs the NIST StRD Linear Regression calculations to within 3 signficant digits		
2.1. Test Linear Regression with dataset Norris, Low difficulty linear		
22.	Type: *Norris < - importData(file = "datasets/Norris.csv")* *Norris.lm.1 < -lm(y~x, data = Norris) summary(Norris.lm.1)*	System responds with: Call: lm(formula = y ~ x, data = Norris) Residuals: Min 1Q Median 3Q Max −2.352 −0.5327 −0.02963 0.6 1.79 Coefficients: Value Std. Error t value Pr(> |t|) (Intercept) −0.2623 0.2328 −1.1267 0.2677 x 1.0021 0.0004 2331.6058 0.0000 Residual standard error: 0.8848 on 34 degrees of freedom Multiple R-squared: 1 F-statistic: 5436000 on 1 and 34 degrees of freedom, the p-value is 0 Correlation of Coefficients: (Intercept) x −0.7738

23.	Type the following: *anova(Norris.lm.1)*	System responds with: Analysis of Variance Table Response: y Terms added sequentially (first to last) Df Sum of Sq Mean Sq F Value Pr(F) x 1 4255954 4255954 5436386 0 Residuals 34 27 1			
24.	Review the information in Figure 2.0a.	Figure 2.0a is the general background information for the NIST StRD Linear Regression test. Initial the INIT column if read and understood. Figure 2.0b is a summary of the data sets referred to in Figure 2.0a.			
25.	Review the information in Figure 2.1a.	Figure 2.1a is the specific model information relevant to the Norris data set. Initial the INIT column if read and understood. Figure 2.1b is the data set information. It is provided for reference.			
26.	Review the results of Step 22 and Step 23 against the certified results of Figure 2.1c.	Results match within 3 significant figures.			
27.	Review the data set in the problem of Figure 2.1d against the data imported in Step 22. (Type Norris at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 2.1d			
2.2.	Test Linear Regression with dataset NoInt1, Average difficulty linear				

28.	Type:
	NoInt1 < - importData(file
	= "datasets/NoInt1.csv")
	NoInt1.lm.1 <-lm(y~–1+x, data
	= NoInt1) summary(NoInt1.lm.1)
		System responds with:
		Call: lm(formula = y ~ –1 + x, data = NoInt1)
		Residuals:
		  Min 1Q Median 3Q Max
		–5.207 –2.521 0.1653 2.851 5.537
		Coefficients:
		    Value Std. Error t value Pr(>|t|)
		x 2.0744 0.0165 125.5000 0.0000			

		Residual standard error: 3.568 on 10 degrees of freedom Multiple R-Squared: 0.9994 F-statistic: 15750 on 1 and 10 degrees of freedom, the p-value is 0

29.	Type the following:
	anova(NoInt1.lm.1)
		System responds with:
		Analysis of Variance Table
		Response: y
		Terms added sequentially (first to last)
		
		`            Df  Sum of Sq   Mean Sq   F Value   Pr(F)`
		`x            1   200457.7  200457.7  15750.25    0`
		`Residuals   10      127.3      12.7`

30.	Review the information in Figure 2.2a.	Figure 2.2a is the specific model information relevant to the NoInt1 data set. Initial the INIT column if read and understood. Figure 2.2b is the dataset information. It is provided for reference.
31.	Review the results of Step 28 and Step 29 against the certified results of Figure 2.2c.	Results match within 3 significant figures.
32.	Review the data set in the problem summary of Figure 2.2d against the data imported in Step 28. (Type NoInt1 at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 2.2d
2.3. Test Linear Regression with data set Filip, High difficulty polynomial		

33.	Type: *Filip <- importData(file = "data sets/Filip.csv")*
	Filip.lm.1 <-lm(y~poly(x,10), data = Filip) attach(Filip)
	poly.transform(poly(x,10), Filip.lm.1))
	coef(Filip.lm.1))
		System responds with:
		`   x^0        x^1        x^2        x^3         x^4`
		`-1467.49   -2772.18   -2316.371  -1127.974   -354.4782`
		`   x^5        x^6        x^7        x^8`
		`-75.1242   -10.87532  -1.062215  -0.06701912`
		`   x^9        x^10`
		`-0.002467811   -4.029625e-05`

34.	Type: *summary(Filip.lm.1)*

System responds with:
Call: lm(formula = y ~ poly(x, 10), data = Filip)
Residuals:

Min	1Q	Median	3Q
-0.008804	-0.002176	4.502e-05	0.002029
Max			
0.007096			

Coefficients:

	Value	Std. Error	t value	Pr(> |t|)
(Intercept)	0.8496	0.0004	2297.8525	0.0000
poly(x, 10)1	0.4614	0.0033	137.8103	0.0000
poly(x, 10)2	-0.0868	0.0033	-25.9256	0.0000
poly(x, 10)3	-0.0827	0.0033	-24.6980	0.0000
poly(x, 10)4	0.0967	0.0033	28.8958	0.0000
poly(x, 10)5	0.0175	0.0033	5.2127	0.0000
poly(x, 10)6	-0.0617	0.0033	-18.4251	0.0000
poly(x, 10)7	0.0067	0.0033	1.9912	0.0503
poly(x, 10)8	0.0340	0.0033	10.1625	0.0000
poly(x, 10)9	-0.0155	0.0033	-4.6397	0.0000
poly(x, 10)10	-0.0150	0.0033	-4.4942	0.0000

Residual standard error: 0.003348 on 71 degrees of freedom
Multiple R-Squared: 0.9967
F-statistic: 2162 on 10 and 71 degrees of freedom, the p-value is 0

Correlation of Coefficients
ALL VALUES ARE ZERO.

35.	Type the following: *anova(Filip.lm.1)*

System responds with:
Analysis of Variance Table
Response: y
Terms added sequentially (first to last)

Df	Sum of Sq	Mean Sq	F Value	Pr(F)
poly(x, 10)	10	0.2423916	0.02423916	2162.44	0
Residuals	71	0.0007959	0.00001121		

36.	Type the following: *detach(2,Filip)*	System responds with: NULL
37.	Review the information in Figure 2.3a.	Figure 2.3a is the specific model information relevant to the Filip data set. Initial the INIT column if read and understood. Figure 2.3b is the data set information. It is provided for reference.
38.	Review the results of Step 33, Step 34 and Step 35 against the certified results of Figure 2.3c.	Results match within 3 significant figures.
39.	Review the data set in the problem of Figure 2.3d against the data imported in Step 33. (Type Filip at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 2.3d.

3. Verify that S-Plus 5.1 performs the NIST StRD Nonlinear Regression calculations to within 3 signficant digits

3.1. Test Nonlinear Regression with data set Misra1a, Lower difficulty exponential

40.	Type: *Misra1a < - importData(file = "data sets/Misra1a1.csv"*
	 The following is a single command:
	 Misra1a.nls.1 < -nls(y~b1*(1-exp (-b2*x)),data = Misra1a,start = list (b1 = 250,b2 = 0.0005)) summary (Misra1a.nls.1)
		System responds with:
		 Formula: y ~ b1 * (1 − exp(- b2 * x))
		 Parameters:
		     Value Std. Error t value
		 b1 2.38942e+02 2.70701e+00 88.2680
		 b2 5.50156e-04 7.26687e-06 75.7075
		
		 Residual standard error: 0.101879 on 12 degrees of freedom
		
		 Correlation of Parameter Estimates:
		     b1
		 b2 −0.999

41.	Review the information in Figure 3.0a.	Figure 3.0a is the general background information for the NIST StRD Nonlinear Regression test. Initial the INIT column if read and understood. Figure 3.0b is a summary of the data sets referred to in Figure 3.0a.		
42.	Review the information in Figure 3.1a.	Figure 3.1a is the specific model information relevant to the Misra1a data set. Initial the INIT column if read and understood. Figure 3.1b is the data set information. It is provided for reference.		
43.	Review the results of Step 40 against the certified results of Figure 3.1c.	Results match within 3 significant figures.		
44.	Review the data set in the problem summary of Figure 3.1d against the data imported in Step 40. (Type Misra1a at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 3.1d.		
3.2. Test Nonlinear Regression with data set Kirby2, Average difficulty rational				
45.	Type: *Kirby2 <- importData(file = "data sets/Kirby2.csv")* The following is a single command: *Kirby2.nls.1 <- nls(y~(b1+b2*x+b3*x^2)/ (1+b4*x+b5*x^2),data = Kirby2,start = list(b1 = 1.5,b2 =-0.15,b3 = 0.0025, b4 =-0.0015,b5 = 0.00002)) summary(Kirby2.nls.1)*	System responds with: Formula: $y \sim (b1 + b2 * x + b3 * x^2)/(1 + b4 * x + b5 * x^2)$ Parameters: |	Value	Std. Error	t value
---	---	---	---
b1	1.67448e+00	8.79894e-02	19.0304
b2	-1.39272e-01	4.11818e-03	-33.8189
b3	2.59610e-03	4.18562e-05	62.0042
b4	-1.72421e-03	5.89314e-05	-29.2578
b5	2.16647e-05	2.01297e-07	107.6260 | Residual standard error: 0.163545 on 146 degrees of freedom		

		Correlation of Parameter Estimates: $\quad$ b1 $\quad$ b2 $\quad$ b3 $\quad$ b4 b2 $\ $ −0.896 b3 $\ \ $ 0.803 $\ $ −0.974 b4 $\ \ $ 0.569 $\ $ −0.793 $\ $ 0.903 b5 $\ \ $ 0.847 $\ $ −0.984 $\ $ 0.962 $\ $ 0.756	
46.	Review the information in Figure 3.2a.	Figure 3.2a is the specific model information relevant to the Kirby2 data set. Initial the INIT column if read and understood. Figure 3.2b is the data set information. It is provided for reference.	
47.	Review the results of Step 45 against the certified results of Figure 3.2c.	Results match within 3 significant figures.	
48.	Review the data set in the problem summary of Figure 3.2d against the data imported in Step 45. (Type Kirby2 at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 3.2d.	
3.3. Test Nonlinear Regression with data set MGH09, Higher difficulty rational			
49.	Type: *MGH09 < - importData (file = "data sets/MGH09.csv")* The following is a single command: *MGH09.nls.1 < - nls(y~(b1*(x^2+b2*x))/ (x^2+x*b3+b4),data = MGH09, start = list(b1 = 0.25,b2 = 0.39, b3 = 0.415,b4 = 0.39)) summary (MGH09.nls.1)*	System responds with: $\quad$ Formula: y ~ (b1 * (x^2 + + b2 * x))/ $\quad$ (x^2 + x * b3 + b4) $\quad$ Parameters: $\quad\quad$ Value $\quad\quad$ Std. Error $\quad$ t value $\quad$ b1 $\ $ 0.192800 $\ $ 0.0114361 $\ $ 16.858900 $\quad$ b2 $\ $ 0.191355 $\ $ 0.1963730 $\ $ 0.974443 $\quad$ b3 $\ $ 0.123030 $\ $ 0.0808503 $\ $ 1.521700 $\quad$ b4 $\ $ 0.136101 $\ $ 0.0900403 $\ $ 1.511560	

		Residual standard error: 0.00662792 on 7 degrees of reedom Correlation of Parameter Estimates: b1 b2 b3 b2 −0.7440 b3 0.0885 0.5250 b4 −0.7640 0.9890 0.4400		
50.	Review the information in Figure 3.3a.	Figure 3.3a is the specific model information relevant to the MGH09 data set. Initial the INIT column if read and understood. Figure 3.3b is the data set information. It is provided for reference.		
51.	Review the results of Step 49 against the certified results of Figure 3.3c.	Results match within 3 significant figures.		
52.	Review the data set in the problem summary of Figure 3.3d against the data imported in Step 49. (Type MGH09 at the S-Plus prompt to see the data set.)	The data imported matches the data in Figure 3.3d.		
4. Verify that S-Plus 5.1 performs a General Additive Model with Gaussian error Distribution and identity link problem correctly to 3 significant digits				

53.	Type: *stack < -data.frame(cbind (stack.x,stack.loss))* (note: stack.x and stack.loss are S-Plus built-in data sets) The following is a single command: *names(stack) < - c("AirFlow","waterTemp", "AcidConc","Loss")*	System responds with:
		Call:
		gam(formula = Loss ~ s(AirFlow) + s(waterTemp) + s(AcidConc),
		      data = stack,
		      control = gam.control(bf.maxit = 50))
		Degrees of Freedom: 21 total; 8.00097 Residual
		Residual Deviance: 67.79171		

	The following is a single command: *stack.gam.1 < -gam (Loss ~ s(AirFlow)+s (waterTemp)+s(AcidConc), control = gam.control(bf.maxit = 50),data = stack) stack.gam.1*	

54.	*summary(stack.gam.1)*	System responds with:
		Call: gam(formula = Loss ~ s(AirFlow) + s(waterTemp) +s(AcidConc), data = stack, control = gam.control(bf.maxit = 50))
		Deviance Residuals:
		  Min 1Q Median 3Q Max
		−3.089759 −1.604992 0.2439517 0.876497 1 3.967667
		(Dispersion Parameter for Gaussian family taken to be 8.472936)
		Null Deviance: 2069.238 on 20 degrees of freedom
		Residual Deviance: 67.79171 on 8.00097 degrees of freedom
		Number of Local Scoring Iterations: 1
		DF for Terms and F-values for Nonparametric Effects
		
		           Df Npar Df Npar F Pr(F)
		(Intercept) 1
		s(AirFlow) 1 3 0.934405 0.4676402
		s(waterTemp) 1 3 3.171167 0.0851828
		s(AcidConc) 1 3 0.975555 0.4507614

55.	Review the results of Steps 53 and 54 against the results of Figure 4 (from Ref. 7)	Results match within 3 significant figures.
56.	Type: *q()*	Successful exit from S-Plus.
57.	Type *exit.*	Successful logoff from my_server_name.

Signoffs

(Signoffs for Document Tester and Test Coordinator with printed name and date spaces:)

Tester   (By signing this the tester of this document agrees that the test has been completed and is accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Tester	

Test Coordinator   (By signing this the tester of this document agrees that the test has been completed and is accurate)

Printed Name	Signed Name	Responsible	Date: MM/DD/YYYY
		Test Coordinator	

Linear, Generalized Linear, and Nonlinear Mixed Effects Models

FARKAD EZZET and JOSÉ C. PINHEIRO

4.1 INTRODUCTION

Biopharmaceutical research often involves the collection of repeated measures on experimental units (such as patients or healthy volunteers) in the form of longitudinal data and/or multilevel hierarchical data. Responses collected on the same experimental unit are typically correlated and, as a result, classical modeling methods that assume independent observations do not lead to valid inferences. Mixed effects models, which allow some or all of the parameters to vary with experimental unit through the inclusion of *random effects*, can flexibly account for the within-unit correlation often observed with repeated measures and provide proper inference. This chapter discusses the use of mixed effects models to analyze biopharmaceutical data, more specifically pharmacokinetic (PK) and pharmacodynamic (PD) data. Different types of PK and PD data are considered to illustrate the use of the three most important classes of mixed effects models: linear, nonlinear, and generalized linear.

Linear mixed effects (LME) models express the response variable as a linear function of both the *fixed effects* and the *random effects*, with an additive within-unit error, see Laird and Wase (1) or Searle et al. (2) for a good review of methodology. The frequentist approach to LME models is generally likelihood-based, with *restricted maximum likelihood* (REML) being the preferred method of estimation (3).

Nonlinear mixed effects (NLME) models extend LME models by allowing the response to be expressed as a nonlinear function of the parameters plus a within-unit error term. Much of this work in biopharmaceutical research began in the 1970s, pioneered by Sheiner and Beal (4). Exact likelihood estimation is generally not feasible, as the marginal distribution of the response cannot be expressed in closed form. Approximate likelihood methods are used instead, with different degrees of accuracy and computational intensity having been proposed in the

literature; see Davidian and Giltinan (5) for a good review of some of these methodologies. A more detailed account of the theory and application of LME and NLME models, especially under S-Plus (6) can be found in work by Pinheiro and Bates (7). Research to produce computationally efficient and accurate approximate likelihood methods for NLME models is still quite active.

Generalized linear mixed models (GLMMs) provide another type of extension of LME models aimed at non-Gaussian responses, such as binary and count data. In these models, conditional on the random effects, the responses are assumed independent and with distribution in the exponential family (e.g., binomial and Poisson) (8). As with NLME models, exact likelihood methods are not available for GLMMs because they do not allow closed form expressions for the marginal distribution of the responses. Quasilikelihood (9) and approximate likelihood methods have been proposed instead for these models.

Mixed effects models under a Bayesian framework have been widely studied and used with the use of Markov chain Monte Carlo methods (10). These methods have gained particular popularity as complex problems became easily formulated using the WinBUGS software (11). See Congdon (12) for an extensive coverage of topics and examples and implementation in WinBUGS.

In this chapter we investigate and illustrate the use of LME and NLME models, as well as GLMMs using algorithms implemented in the S-Plus functions lme, nlme, and glme, respectively. We attempt to demonstrate that, even under fairly complex hierarchical, correlated data structures, the existing algorithms are capable of properly estimating the underlying parameters (fixed effects and variance–covariance components), thus providing reliable unbiased inference.

We begin by considering a simple PK dose proportionality (DP) study in which subjects receive an experimental drug to evaluate if the increase in exposure is proportional to dose. We examine the problem in two ways: (a) using an exposure metric, for example, area under the concentration–time curve (AUC), which leads to an LME model; and (b) using the concentration data directly, which requires the use of an NLME model. Concentration data are simulated using different hierarchical random effects structures. We then extend the DP example to include a clinical response and explore a pharmacokinetic/pharmacodynamic (PK/PD) NLME model. Collapsing the clinical response into a binary measure allows the illustration of GLMMs.

Common features among the three different classes of models and their implementation within the S-Plus environment come into light during the analysis of the examples: in particular, the syntax for defining the fixed and random effects in the models, as well as methods for extracting estimates from fitted objects. All data sets discussed in this chapter are fictitious: that is, they are generated by simulation. The reader is encouraged to experiment with the code provided in Appendix 4.1 to explore alternative scenarios.

4.2   PHARMACOKINETIC DOSE PROPORTIONALITY PROBLEM

Consider a dose proportionality study in which each subject is to receive a number of doses, usually two or more, of an experimental drug to evaluate if exposure increases proportionally with dose. We adopt a crossover design and, to keep things simple, assume that issues related to carryover, period, and sequence effects, as well

as subject by dose interactions (13), are of no concern in this example. The S-Plus function `sim.dp.mult`, included in Appendix 4.1, generates drug concentrations (C) at times (t) following drug administration, C(t), according to the single dose oral one-compartment PK model:

$$C(t) = \frac{F \cdot dose \cdot K_a}{V(K_a - K_e)} [e^{-k_e t} - e^{-k_a t}]$$ (4.1)

Typically, two types of error are recognized: (a) measurement level error resulting from error in concentrations due to assays, time of measurement, and so on and (b) subject level error represented in the model by random effects, accounting for deviations in the PK parameters between subjects, that is, in absorption (K_a), elimination (K_e), and/or volume (V). V is usually expressed as V/F when the fraction of dose absorbed (F) is unknown.

Formally, we may express C(t) as $C_{ij}(t) = f(\theta_i, dose_j, t) [1 + \varepsilon_{ij}(t)]$, where $C_{ij}(t)$ and $f(\theta_i, dose_j, t)$ are the measured and predicted concentrations for the ith subject at the jth dose at time t, respectively, and θ_i is the vector of PK parameters for the ith subject. Here, the intersubject variability in the PK parameters is assumed proportional. For instance, volume for the ith subject is defined as $V_i = V \exp(\eta_{i,V})$, where the random effects $\eta_{i,V}$ are independently distributed as $N(0, cv \cdot V)$. The prefix cv denotes coefficient of variation for V. The measurement error is assumed multiplicative, with the $\varepsilon_{ij}(t)$ independently distributed as $N(0, cv \cdot \varepsilon)$. The functional form of f is determined by the type of PK model being considered; for the DP example we assume the one-compartment model described above.

A third possible source of variation, accounting for deviations in the PK parameters within subject from period to period, often referred to as interoccasion (IO) variability, may also be incorporated in the PK model. For example, we may define $V_{ij} = V \exp(\eta_{i,V} + \eta'_{ij,V})$, where $\eta_{i,V}$ as before represents the intersubject random effect while $\eta'_{ij,V}$ represents the interoccasion random effect within subject, assumed independently distributed as $N(0, cv \cdot occ \cdot V)$.

The S-Plus data frame `dp1` is generated by calling the function `sim.dp.mult` assuming strict dose proportionality and no IO variability, as illustrated below. Figure 4.1 shows a trellis display of the corresponding concentration–time profiles.

```
dose <- c(5, 10, 20)
time <- c(0.5,1,2,4,6,8,12,24,36,48,72,96)
dp1 <- sim.dp.mult(nsub = 12, Pars = c(ka = .1, ke = .03, v = 4),
    cv.sub = c(ka = .3, ke = .3, v = .3),
    cv.error = 0.1, time = time, dose = dose, seed = 123)
```

A plot of the observed *AUC* (calculated using trapezoidal rule implemented in the function `aucTrap`) versus dose (not shown here) reveals an almost perfect linear relationship, a consequence of the strict dose proportionality used to simulate the data.

4.2.1 DP Using *AUC* in a Linear Mixed Effects Model

Conventional DP analysis proceeds with the calculation of *AUC* using the trapezoidal rule, followed by an analysis of variance of the resulting values normalized
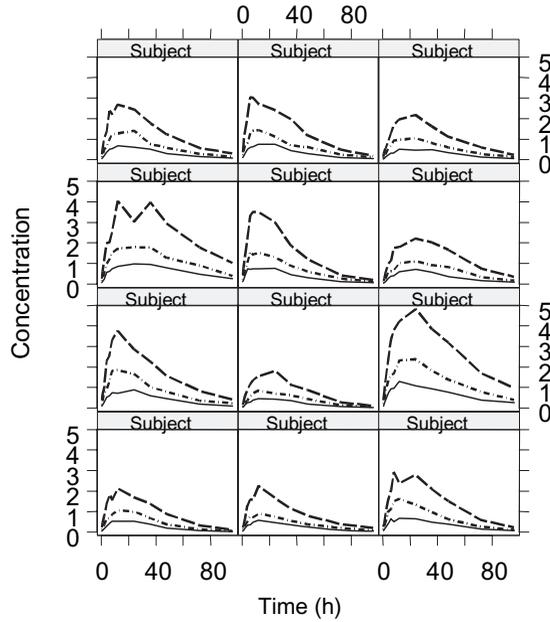
FIGURE 4.1   Concentration–time profiles generated using `sim.dp.mult`.

by dose. A more appealing approach (14) is to define AUC as a function of dose using a power model:

$$AUC_{ij} = a(dose_j)^{\beta}v_{ij} \qquad (4.2)$$

with v_{ij} representing an error term. Applying log to both sides gives a linear model with a DP parameter β.

$$\log(AUC_{ij}) = \log(a) + \beta \log(dose_j) + \log(v_{ij}) = \alpha + \beta \log(dose_j) + \varepsilon_{ij} \qquad (4.3)$$

Strict DP is achieved when $\beta = 1$. Accounting for within-subject correlation due to repeated measures on the same subject is accomplished by introducing a subject effect S:

$$\log(AUC_{ij}) = \alpha + \beta \log(dose_j) + S_j + \varepsilon_{ij} = \alpha_i + \beta \log(dose_j) + \varepsilon_{ij} \qquad (4.4)$$

where $\alpha_i = \alpha + S_j$ represents the subject intercept. The error ε_{ij} combines measurement error in AUC and any other sources of error, including model misspecification. If α_i is treated as a fixed effect, standard linear regression analysis for independent data can be used, but information about intersubject variation cannot be provided. Instead, we consider a linear mixed effects model in which α_i is assumed random. This is done using the `lme` function in S-Plus with the following function call using data set *dpauc1*:

```
f.dpauc1 <- lme(log(auc) ~ log(dose), data=dpauc1, ~1|subject)
```

resulting in the following parameter estimates:

```
Random effects:
Formula: ~ 1 | subject
  (Intercept)  Residual
StdDev: 0.3650006 0.04121098
Fixed effects: log(auc) ~ log(dose)
               Value Std.Error DF t-value p-value
(Intercept) 1.940150 0.1092254 23 17.76281 <.0001
  log(dose) 0.981528 0.0121362 23 80.87623 <.0001
```

The random intercept has an estimated standard deviation (SD) of 0.37 or, equivalently, a coefficient of variation of 37% in the original AUC scale, reflecting the combined subject variation in all three PK parameters—K_a, K_e, and V. The estimated SD of measurement error in $\log(AUC)$ is small, 0.04. The DP parameter (i.e., the coefficient of $\log(AUC)$) is estimated at 0.98, with a 95% confidence interval (CI) of (0.96,1.01), consistent with strict dose proportionality.

4.2.2 DP Using Concentration Data in a Nonlinear Mixed Effects Model

We may alternatively tackle the DP problem by analyzing the raw concentration data directly, using a reparameterization of the assumed PK model. Because the PK model is nonlinear in its parameters, an NLME model is needed. In this example, since interest is centered on the DP assumption, we redefine the oral dose one-compartment model to have AUC as one of its parameters, using the relation $k_e = dose/(AUC \times V)$. The function comp1.oral.auc.log defined below implements the reparameterized oral dose one-compartment model. To enforce positive estimates of the PK parameters, we have chosen to estimate the parameters on the log scale, hence the prefix (l = log) preceding the PK parameter names in the argument list of the function.

```
comp1.oral.auc.log = function(lka, lauc, lv, f, dose, time)
{
  ka = exp(lka); auc = exp(lauc); v = exp(lv); ke = dose/(auc * v)
  (ka * dose * f)/(v * (ka - ke)) * (exp( - ke * time) - exp( - ka * time))
}
```

The nlme call for fitting the corresponding nonlinear mixed effects model is

```
f.dp1b <- nlme(conc ~ comp1.oral.auc.log(lKa, lAUC, lV, 1, dose,
time),
         data=dp1, fixed = list(lKa + lV ~ 1, lAUC ~ log(dose)),
         random = list(subject = pdBlocked(list(lKa ~ 1, lAUC +
               lV ~ 1))),
         weights = varPower(), start = coef(f.dp1a), verbose = T)
```

Starting values for fixed effects in the above function call were extracted from a fit object (`f.dp1a`) using the gnls function, a nonlinear fitting method using generalized least squares. This is optional as the user may provide his/her own list of starting values.

We assume a block-diagonal matrix of subject random effects such that $\eta_{j,lAUC}$ and $\eta_{j,lV}$ are correlated, but independent of $\eta_{j,lka}$. Here are the resulting estimates:

```
Random effects:
  Composite Structure: Blocked
  Block 1: lKa
  Formula: lKa ~ 1 | subject
          lKa
StdDev: 0.3078118
    Block 2: lAUC, lV
  Formula: list(lAUC ~ 1 , lV ~ 1 )
  Level: subject
  Structure: General positive-definite
            StdDev Corr
    lAUC 0.3869677 lAUC
      lV 0.2228085 -0.925
Residual 0.1006868

Variance function:
  Structure: Power of variance covariate
  Formula: ~ fitted(.)
  Parameter estimates:
    power
  0.9985222
Fixed effects: list(lKa + lV ~ 1, lAUC ~ log(dose))
                      Value    Std.Error DF   t-value   p-value
              lKa -2.266270   0.0918522 417  -24.6730  <.0001
               lV  1.416695   0.0669196 417   21.1701  <.0001
lAUC.(Intercept)   1.964468   0.1138468 417   17.2554  <.0001
  lAUC.log(dose)   0.998404   0.0078681 417  126.8926  <.0001
```

There is a close similarity between the estimates for the intercept and slope parameter corresponding to $\log(AUC)$ in the `lme` and `nlme` fits. It is not surprising that good estimates are obtained in this data-rich situation. However, reducing the number of time points in `dp1` to include only measurements at 0.5, 1, 4, 12, and 48 hours results in similar estimates, though with a larger SE, suggesting the ability of performing a DP study with a sparse sampling scheme.

```
                   Value Std.Error DF t-value p-value
lAUC.(Intercept) 2.001881 0.1258471 165 15.90724 <.0001
  lAUC.log(dose) 0.987246 0.0241469 165 40.88503 <.0001
```

Thus, in the case of sparse blood sampling schedules using concentration data in an NLME model offers a practical alternative to using (inaccurately) calculated

AUC values in an LME model. With the latter approach, an extensive schedule is needed for accurate *AUC* determination, in addition to a reliable method for capturing the *AUC* portion between the last quantifiable concentration value and infinity.

4.2.3 DP Using Concentration Data in a NLME Model with Multilevel Random Effects (Interoccasion Variability)

The LME model of Section 4.2.1 and the NLME model of Section 4.2.2 both involved two random components: measurement error and subject random effects. In this section we explore a two-level random effect hierarchy by introducing IO variability in the PK parameters (K_a, K_e, and V), so that the subject's parameters may vary from period to period. Note that this is not a period effect, but rather an uncontrollable random variation in the subject's pharmacokinetics. The data frame `dp2`, incorporating IO random effects, is obtained by calling `sim.dp.mult` as follows:

```
dp2 <- sim.dp.mult(nsub=12, Pars =c(ka=.1, ke=.03, v=4),
                   cv.sub = c(ka=.3, ke=.3, v=.3),
                   cv.occ = c(ka=0.2, ke=0.2, v=0.2),
                   cv.error=0.1, time = time, dose = dose)
```

The magnitude of change in K_e, K_a, or V due to IO is set to have a *cv* equal to 20%. A plot of calculated *AUC* versus dose (not shown) reveals the influence of IO variability on *AUC*, leading to a nonlinear relationship with dose.

Ignoring IO variability and calling `lme` as before gives the following estimates:

```
Random effects:
  Formula: ~ 1 | subject
          (Intercept) Residual
StdDev: 0.3779992 0.182379

Fixed effects: log(auc) ~ log(dose)
               Value Std.Error DF t-value p-value
(Intercept) 2.080637 0.1677046 23 12.40656 <.0001
  log(dose) 0.915968 0.0537086 23 17.05440 <.0001
```

Because *AUC* is a function of K_e and *V*, its variance is increased with IO variability, impacting the measurement error variability in the `lme` fit above—hence the larger residual standard deviation as compared with that obtained in the `lme` fit of the `dp1` data. Alternatively, we resort to using the raw concentration data and incorporate IO random effects in the NLME model by allowing the parameters K_a, *AUC*, and *V* to vary between dose administrations. This is implemented in the `nlme` call using the following `random` statement:

```
random = list(subject = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1)),
                 dose = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1)))
```

The second component of the list represents the IO random effects associated with different dose administrations. The fixed effects estimates for the above model are

```
                  Value Std.Error DF  t-value   p-value
          lKa -2.233085 0.1110798 393 -20.10343 <.0001
           lV  1.427694 0.0843185 393  16.93215 <.0001
lAUC.(Intercept) 1.985077 0.1572971 393  12.61992 <.0001
  lAUC.log(dose) 0.986582 0.0459793 393  21.45710 <.0001
```

and the variance–covariance parameter estimates are

```
  Block 1: lKa
Formula: lKa ~ 1 | subject
             lKa
StdDev: 0.3574699

  Block 2: lAUC, lV
Formula: list(lAUC ~ 1 , lV ~ 1 )
Level: subject
Structure: General positive-definite
      StdDev Corr
lAUC 0.3853345 lAUC
  lV 0.2646356 -0.904

Composite Structure: Blocked
  Block 1: lKa
Formula: lKa ~ 1 | dose %in% subject
             lKa
StdDev: 0.1968827

  Block 2: lAUC, lV
Formula: list(lAUC ~ 1 , lV ~ 1 )
Level: dose %in% subject
Structure: General positive-definite
          StdDev Corr
   lAUC 0.1903516 lAUC
     lV 0.1785101 -0.631
Residual 0.1014465

Variance function:
  Structure: Power of variance covariate
  Formula: ~ fitted(.)
  Parameter estimates:
     power
  0.9855959
```

Notice that appropriately adjusting for IO variability in the NLME model not only reduced the bias in the estimate of the dose proportionality parameter but also led to a valid estimate of the residual error standard deviation, a value much closer to that used to simulate the data, 0.1.

4.2.4 Comparing Sample and Model Estimates of Random Effects Parameters

Although the fixed effects have been well estimated, it is also of interest to examine how closely the estimated standard deviations of the random effects reflect the true variability in the simulated data. The `dp2` data frame includes values of the generated subject random effects, interoccasion random effects, and measurement errors, from which *sample* variances can be obtained and compared to the model estimates. The intersubject sample standard deviations of $\log(K_a)$, $\log(AUC)$, and $\log(V)$ are 0.33, 0.41, and 0.23, respectively. The corresponding model estimates are 0.36, 0.39, and 0.26. For the IO random effects, the sample SDs are 0.17, 0.22, and 0.17, while the corresponding values obtained in the model fit are 0.20, 0.19, and 0.18, respectively. The sample and model *SD* for measurement error are both equal to 0.1, indicating a good agreement overall between sample and model estimates.

4.3 PHARMACOKINETIC–PHARMACODYNAMIC (PK-PD) MODEL

The function `sim.pkpd.mult` simulates a clinical response (R) on the basis of a PK/PD model. It incorporates a combined placebo (P) effect and a drug (D) effect. The placebo effect at time t is defined as

$$P(t) = BL \{1 - \exp(-a_1 t) + \exp(-a_2 t)\}, \, a_1 > a_2 > 0$$

while the drug effect is defined as

$$D(t) = (E_{max} C(t))/(EC_{50} + C(t)) \text{ and } R(t) = \{P(t) + D(t)\} (1 + \varepsilon(t))$$

The placebo model assumes an endogenous response, influenced by baseline (BL), and two exponential functions. For $a_1 > a_2$, $P(t)$ increases over time above BL then declines back to BL for sufficiently large t. The drug model is a stimulus model, a function of plasma concentration $C(t)$, maximal effect E_{max}, and EC_{50}, the concentration that produces 50% of the maximal effect. In this example, $C(t)$ is generated without measurement error but is influenced by subject random effects in K_a, K_e, and V, as discussed in the previous section. Additional subject random effects are considered for BL and E_{max}.

The data frame `pkpd1` is generated according to the PK-PD model above, using the following call to the function `sim.pkpd.mult`. Figure 4.2 shows a trellis display of the corresponding time profiles for the simulated PD response.

```
time   <- c(0, 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 14, 21, 28) # days
time   <- time * 24 # convert to hours
pkpd1 <- sim.pkpd.mult(nsub=24, doseint=24, ndose=29,
        Pars =c(ka=.1, ke=.03, v=4, bl=8, ec50=5, emax=8),
        pdPars=c(a1=0.1, a2=0.05),
        cv.sub = c(ka=.3, ke=.3, v=.3, bl=0.35,
                   ec50=0.0, emax=0.35),
        cv.error = 0.05, time = time, dose = c(10),
        levIncCV = 0, parsForm = list(ke = ~ke*sqrt(dose/dose)),
        seed = 123)
```
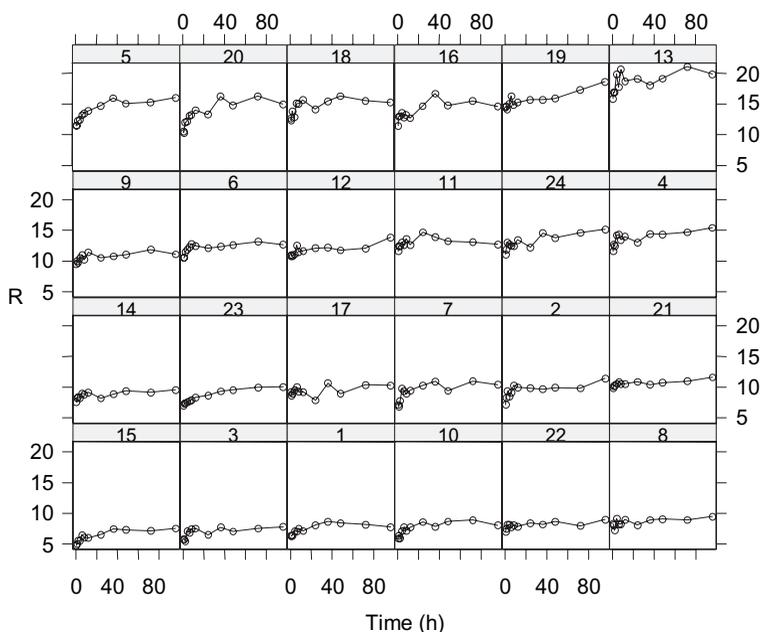
FIGURE 4.2   Response–time profiles generated using `sim.pkpd.mult`.

The `nlme` function was used to fit the PK/PD model using the following function call:

```
f.pd1 <- nlme(resp ~ placebo.log(lbl, la1, la2, time)
               + drug.log(lemax, lec50, concm),
  data=pkpd1, fixed = list(lbl+la1+la2+lemax+ lec50~1),
       random  = list(subject = pdDiag(list(lbl~1, lemax ~ 1))),
          start = log(c(15, .2, .1, 15, 10)),
          weight = varPower(), verbose = T)
```

The estimation results for the `f.pd1` fit are as follows:

```
Random effects:
  Formula: list(lbl ~ 1 , lemax ~ 1 )
  Level: subject
  Structure: Diagonal
               lbl lemax Residual
StdDev: 0.3244805 0.2962879 0.04937675

Variance function:
  Structure: Power of variance covariate
  Formula: ~ fitted(.)
  Parameter estimates:
     power
  0.9912175
```

```
Fixed effects: list(lbl + la1 + la2 + lemax + lec50 ~ 1)
          Value  Std.Error  DF  t-value  p-value
  lbl   2.186591  0.067473  284  32.40690  <.0001
  la1  -2.722665  1.174003  284  -2.31913  0.0211
  la2  -2.749036  1.176724  284  -2.33618  0.0202
lemax   1.970194  0.255620  284   7.70751  <.0001
lec50   1.433258  0.406105  284   3.52928  0.0005
```

Using `exp(fixef(f.pd1))` produces estimates of fixed effects, in agreement with the values used in the simulation (i.e., bl = 8, a1 = 0.1, a2 = 0.05, emax = 8, ec50 = 5).

```
exp(fixef(f.pd1))
     lbl la1       la2         lemax      lec50
8.904805 0.0656994 0.06398949 7.172067 4.192337
```

The remaining parameters, representing the variance and covariance components, are also fairly accurately estimated in this example. All confidence intervals for the model parameters contain the corresponding value used to simulate the data.

4.4 REPEATED BINARY MEASURE: GLMM FIT

To illustrate the use of GLMMs to analyze biopharmaceutical data, we artificially added binary response variable $R_b(t)$ to the `pkpd1` data of the previous section by creating an indicator variable for the event that the *PD* response was >12.

```
pkpd1$Rb <- as.integer(pkpd1$resp > 12)
```

The `glme` function implements GLMMs in S-Plus, being available in the experimental library `S+CorrelatedData`, which can be downloaded from the Insightful Corporation website at `www.insightful.com` (it requires Version 6.2 or higher of S-Plus). Its syntax is almost identical to that of `lme`, with an additional argument —`family`, representing the desired exponential family distribution to be used. Most commonly used families are `binomial` and `poisson`, for binary and count data, respectively.

For example, to fit a model with a single mean parameter and a single subject random effect, one could use

```
f.bin1 <- glme(Rb ~ 1, pkpd1, ~1|subject, family = binomial)
```

producing the following estimation results:

```
Generalized linear mixed-effects model fit by restricted PQL
Family: Binomial with Logit link
. . .
```

```
Random effects:
  Formula: ~ 1 | subject
        (Intercept) Residual
StdDev:  4.328903   0.6263724

Fixed effects: Rb ~ 1
                Value Std.Error DF  t-value   p-value
(Intercept) -0.955206 0.9430668 288 -1.012872 0.312
```

The interpretation of the estimates in relation to those of the PK/PD model is not straightforward. The intercept estimate gives the logit of the probability that the PD response is >12, with the random effect SD corresponding also to the logit scale. The default estimation method used in `glme` is restricted penalized quasilikelihood (PQL), (9). The question of primary interest is factors influencing the dichotomized response variable R_b. Here, it is a question of whether its probability of taking the value 1 (i.e., of the PD response being >12) changes with drug concentration. We can investigate that by fitting a different GLMM,

```
f.bin2 <- glme(Rb ~ concm, pkpd1, ~ 1 | subject, family =
binomial)
```

with estimated fixed effects:

```
Fixed effects: Rb ~ concm
                Value Std.Error DF  t-value  p-value
(Intercept) -6.756281  1.679655  287 -4.02242 0.0001
     concm   2.770181  0.263541  287 10.51138 <.0001
```

The highly significant and positive estimate for the concentration (`concm`) slope indicates that the logit of the probability increases with increasing concentration, or more precisely, that the probability that the PD response exceeds 12 increases when the concentration increases.

4.5   MODEL UNCERTAINTY: SIMULATION

With satisfactory model diagnostics, we may wish to evaluate model predictions. Predictions are useful in a number of ways, notably for evaluating model behavior under alternative settings, for example, a range of dosage regimens, or to establish the likely individual response to specific study design features. We thus distinguish two types of predictions: population and individual.

Since *population predictions* represent mean estimates, these are determined using model estimates (μ) and its variance–covariance matrix (Σ), characterizing uncertainty in the model estimates. Thus, a single population profile is obtained by making a single draw from a multivariate normal (MVN) with mean μ and variance Σ, and substituting in the model equation. The estimates μ and Σ can be read using `fixef(obj)` and `obj$varFix`, respectively, where `obj` is the name of the fitted object. However, we combined these two steps using the function

`simPars(getFixPars(obj), N = 1)`. As an illustration, we generate $N = 12$ population profiles based on the model fit f.pd1.b at a 20 mg dose. Figure 4.3 provides the concentration–time profile while Figure 4.4 provides the response–time profile.

Individual predictions, however, require in addition to μ and Σ the estimated standard deviations of the random subject effects (τ) and corresponding variance–covariance matrix (Ω). The extraction of τ and Ω is less straightforward; but the function `getRanPars` does just that. Notice that this function also extracts estimates and standard errors of other random effects parameters, for example,
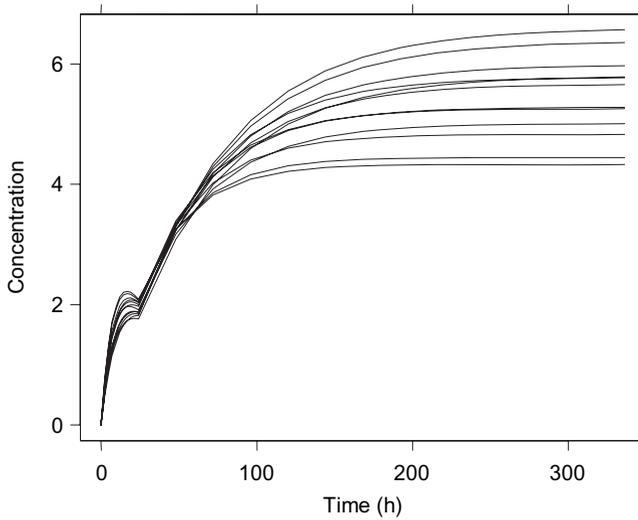
FIGURE 4.3   Population concentration–time profiles (12 replicates).
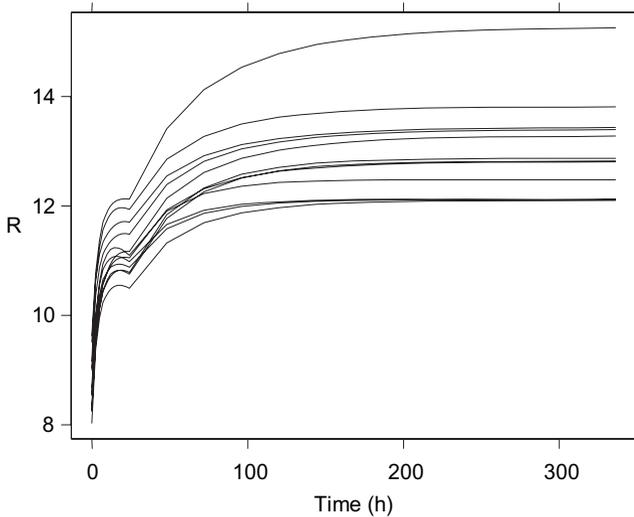
FIGURE 4.4   Population response–time profiles (12 replicates).
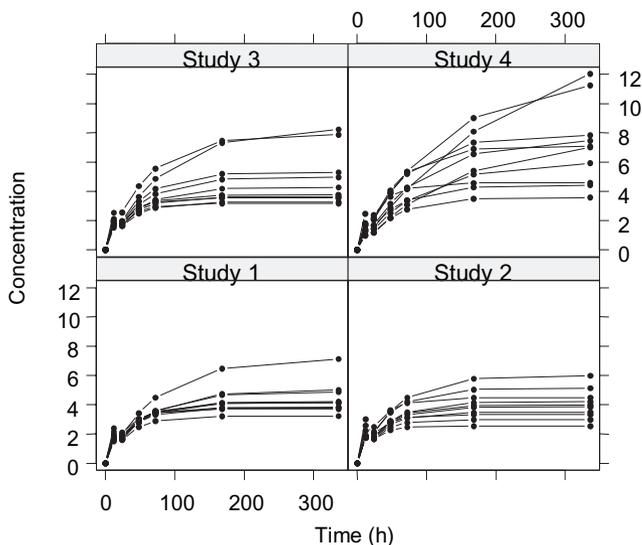
FIGURE 4.5 Individual concentration–time profiles ($N = 12$) from four replicated studies.

measurement error, or correlation coefficients if random effects are assumed correlated. Therefore, caution should be exercised when extracting the appropriate elements, by appropriately matching the function output with that of the model fit. A random draw of an individual subject random effects can be obtained using `simPars(getRanPars(obj), N = 1)`. Substituting into the model equation of a population profile provides an individual profile from that population. As an illustration, we generate $N = 12$ individual profiles based on four different population realizations (representing four study results) using the model fit f.pd1.b at a 20 mg dose. Figure 4.5 shows the concentration–time profile while Figure 4.6 shows the response–time profile in trellis plots.

Notice the range of resulting clinical response in the four panels of Figure 4.6, reflecting uncertainty in the estimates as well as magnitude of intersubject variance. The process described above may be used to calculate various statistics. This can be particularly useful at the design stage of future clinical trials. For instance, we may be interested in computing the 5th and 95th percentiles of R at day 7 of treatment under two sample size scenarios, say, 24 or 36 subjects. For each design, percentiles are determined based on a large number (say, 1000) of replicated studies. Figure 4.7 depicts the distributions of 5th and 95th percentiles under the two designs, suggesting some, although no substantial, gain in precision is achieved with the larger sample size.

4.6   SUMMARY

In this chapter we introduce and illustrate the use of linear, nonlinear, and generalized linear mixed effects models within the S-Plus environment. Based on personal experience, the fitting algorithms implemented in these S-Plus functions appear
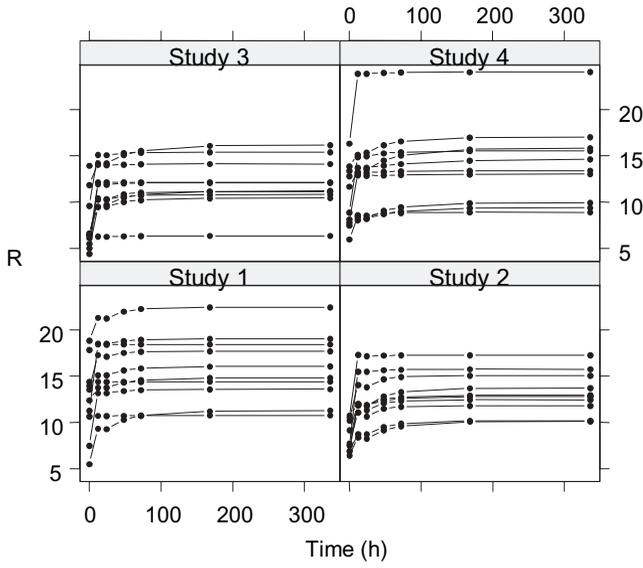
FIGURE 4.6   Individual response–time profiles ($N = 12$) from four replicated studies.
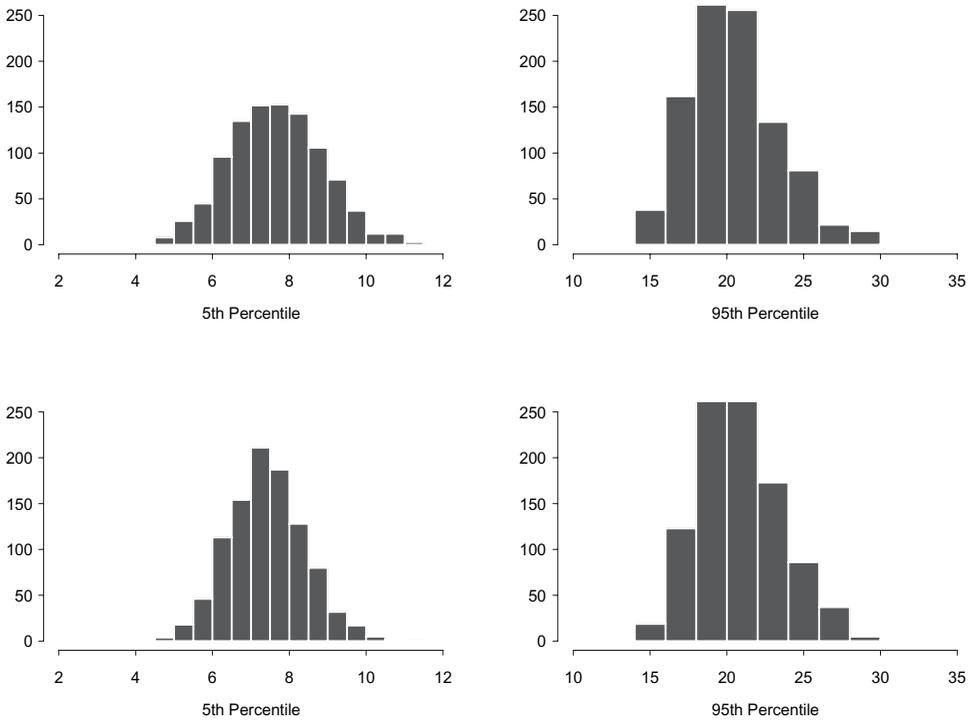
FIGURE 4.7   Histograms of response's 5th and 95th percentiles. *Top*: $N = 24$; *bottom*: $N = 36$.

to be stable and reliable, producing nearly unbiased estimates for problems with appropriate sample sizes. The syntax of function calls and extraction methods of results are similar among the various functions, rendering the environment convenient for a wide range of models including multilevel hierarchical variance models. One limitation of the `nlme` function in S-Plus is for applications in which the model function is not expressed in closed form (e.g., a system of differential equations). Recently, researchers have extended the NLME software to allow a link between the `nlme` function and a linear differential equation solver, which can be used to fit more complex NLME models. The extended `nlme` function is currently only available in the R language, as part of the nlmeODE contributed library (`http://nlmeode.sourceforge.net`).

We investigate a simple dose proportionality (DP) problem that can be successfully analyzed with software for linear regression analysis with independent data. The DP problem is expanded to include random effects and is thus dealt with as a linear mixed effects problem using a summary PK measure (AUC), or as a nonlinear mixed effects problem (up to two levels of random effects) using raw concentration data. We tend to favor the latter approach, especially in cases of sparse concentration data. This may be particularly true in studies involving special populations (e.g., pediatrics) or when using patient data from Phase 2 or Phase 3 studies. These studies offer the opportunity to reevaluate assumptions of dose proportionality, bioequivalence, and drug interactions in a larger, more representative, patient population under varied clinical settings.

The exposition is restricted to two levels of random effects, which are illustrated as subject random effects and interoccasion variability. Another similar situation with two levels of random effects may involve random subject effects nested within random center or study effects. In preclinical allometric studies, used to predict human drug exposure from animal studies, animal species can be considered as a random effect in which each species deviates from an allometry relationship (often drug clearance and body weight) by a fixed but unknown amount. Although the methods discussed in this chapter are able to deal with more than two levels of grouping in the data, the ability to reliably estimate a larger number of variance components is determined by the availability of data at each level. This has not been explored herein; thus, no guidance is offered on appropriate sample sizes and the stability of estimation algorithms under these models.

When using mixed effects models in practice, of particular importance is the derivation of predictions following the model fit. The fixed effect estimates and their corresponding standard errors allow calculation of population (or mean) predictions and associated uncertainty intervals. The estimated covariance matrix for the random effects together with the corresponding standard errors allow calculation of individual predictions and uncertainty intervals. These methods are particularly helpful at the design stage of new studies, when combined with modeling and simulation approaches.

We have included in Appendix 4.1 the S-Plus functions and scripts used in the simulation and analysis of the examples presented here. The reader should be able to reproduce the results using the same seed (=123) and is encouraged to attempt variations to explore other possibilities. The S-Plus code was not developed with computational efficiency in mind, but just to illustrate how the different mixed effects models can be used to analyze PK and PD data in S-Plus. We hope we have

provided a brief account of mixed effects models and a framework for exploring such models under S-Plus.

REFERENCES

1. N. M. Laird and J. H. Ware, Random-effects models for longitudinal data. *Biometrics* **38**:963–974 (1982).

2. S. R. Searle, G. Casella, and C. E. McCulloch, *Variance Component*, Wiley, Hoboken, NJ, 1992.

3. D. A. Harville, Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* **72**:320–340 (1977).

4. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis–Menten model: routine clinical pharmacokinetic data. *J Pharmacokin Biopharm* **8**:553–571 (1980).

5. M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London, 1995.

6. S-Plus, Insightful Corporation, `www.insightful.com`.

7. J. C. Pinheiro and D. M. Bates, *Mixed Effects Models in S and S-PLUS*. Springer-Verlag, New York, 2000.

8. K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models. *Biometrika* **73**:13–22 (1986).

9. N. E. Breslow and D. G. Clayton, Approximate inference in generalized linear mixed models. *J Am Stat Assoc* **88**:9–25 (1993).

10. W. R. Wilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, 1996.

11. WinBUGS software, `www.mrc-bsu.cam.ac.uk/bugs`.

12. P. Congdon, *Applied Bayesian Modelling*. Wiley, Chichester, 2003.

13. S. Senn, *Cross-over Trials in Clinical Research*. Wiley, Chichester, 2002.

14. S. Senn and F. Ezzet, Clinical cross-over trials in phase I. *Stat Methods Med Res* **8**:263–278 (1999).

APPENDIX 4.1   S-PLUS CODE

```
## Scripts used to produce data sets and perform analysis
## Modeling and simulation functions included at the end of this
## file should be sourced into S-Plus before the scripts can be run

## Dose proportionality example of Section 4.2
dose <- c(5, 10, 20)
time <- c(0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48, 72, 96)
dp1 <- sim.dp.mult(nsub=12, Pars =c(ka=.1, ke=.03, v=4),
              cv.sub = c(ka=.3, ke=.3, v=.3),
              cv.error = 0.1, time = time, dose = dose, seed = 123)
```

```
dp1.Gd<- groupedData(conc ~ time | subject/dose, dp1,
                labels = list(x = "Time", y = "Concentration"),
                units = list(x = "(hrs)"))


trellis.device()
plot(dp1.Gd, display = 1, collapse = 1, inner = ~dose, aspect = 1, grid=F)


auc1 <- aucTrap(dp1)
sub <- unique(dp1$subject)
doses <- unique(dp1$dose)
dpauc <- expand.grid(dose=doses,subject=sub)
dpauc1 <- cbind(expand.grid(dose=doses,subject=sub), auc = auc1)
dpauc1 <- groupedData(auc ~ dose|subject, dpauc1,
                    labels = list(x = "Dose", y = "AUC"))


## lme fit of Section 2.1
f.dpauc1 <- lme(log(auc) ~ log(dose), data=dpauc1, ~1|subject)
summary(f.dpauc1)
intervals(f.dpauc1)


## nlme fit of Section 4.2.2
f.dp1a <- gnls(conc ~ comp1.oral.auc.log(lKa, lAUC, lV, 1, dose, time),
        dp1, params = list(lKa + lV ~ 1, lAUC ~ log(dose)),
        start = log(c(0.1, 100, 1, 4)), verbose = T)


f.dp1b <- nlme(conc ~ comp1.oral.auc.log(lKa, lAUC, lV, 1, dose, time),
        data=dp1, fixed = list(lKa + lV ~ 1, lAUC ~ log(dose)),
        random = list(subject = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1))),
        weights = varPower(), start = coef(f.dp1a), verbose = T)
summary(f.dp1b)
intervals(f.dp1b)


dp1sub <- dp1[is.element(dp1$time, c(0.5, 1, 4, 12, 48)),]
f.dp1bSub<-nlme(conc~comp1.oral.auc.log(lKa,lAUC,lV,1,dose,time),
        data=dp1sub, fixed = list(lKa + lV ~ 1, lAUC ~ log(dose)),
        random = list(subject = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1))),
        weights = varPower(), start = coef(f.dp1a), verbose = T)
summary(f.dp1bSub)
intervals(f.dp1b)


## IO example of Section 4.2.3
dp2 <- sim.dp.mult(nsub=12, Pars =c(ka=.1, ke=.03, v=4),
             cv.sub = c(ka=.3, ke=.3, v=.3),
             cv.occ = c(ka=0.2, ke=0.2, v=0.2),
             cv.error=0.1, time = time, dose = dose, seed = 123)
```

```
auc2 <- aucTrap(dp2)
sub <- unique(dp2$subject)
doses <- unique(dp2$dose)
dpauc <- expand.grid(dose=doses,subject=sub)
dpauc2 <- cbind(expand.grid(dose=doses,subject=sub), auc = auc2)
dpauc2 <- groupedData(auc ~ dose|subject/dose, dpauc2,
             labels = list(x = "Dose", y = "AUC"))


## lme fit
f.dpauc2 <- lme(log(auc) ~ log(dose), data=dpauc2, ~1|subject)
summary(f.dpauc2)
intervals(f.dpauc2)


# nlme fits
f.dp2a <- gnls(conc ~ comp1.oral.auc.log(lKa, lAUC, lV, 1, dose, time),
        dp2, params = list(lKa + lV ~ 1, lAUC ~ log(dose)),
        start = log(c(0.1, 100, 1, 4)), verbose = T)


f.dp2b <- nlme(conc ~ comp1.oral.auc.log(lKa, lAUC, lV, 1, dose, time),
        data=dp2, fixed = list(lKa + lV ~ 1, lAUC ~ log(dose)),
        random = list(subject = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1))),
        weights = varPower(), start = coef(f.dp2a), verbose = T)


f.dp2c <- update(f.dp2b, random = list(subject = pdDiag(lKa + lAUC + lV ~ 1),
        dose = pdDiag(lKa + lAUC + lV ~ 1)))


f.dp2d <- update(f.dp2c, start = list(random = ranef(f.dp2c)),
        random = list(subject = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1)),
        dose = pdBlocked(list(lKa ~ 1, lAUC + lV ~ 1))))


summary(f.dp2d)
intervals(f.dp2d)


## PK-PD model example of Section 4.3
time <- c(0,.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 14, 21, 28) # days
time <- time * 24                                # convert to hours


pkpd1 <-
  sim.pkpd.mult(nsub=24, doseint=24, ndose=29,
        Pars =c(ka=.1, ke=.03, v=4, bl=8, ec50=5, emax=8),
        pdPars=c(a1=0.1, a2=0.05),
        cv.sub = c(ka=.3, ke=.3, v=.3, bl=0.35, ec50=0.0,
          emax=0.35),
        cv.error = 0.05, time = time, dose = c(10),
        levIncCV = 0, parsForm = list(ke = ~ke*sqrt(dose/dose)),
        seed = 123)
```

```
pkpd1.Gd<- groupedData(resp ~ time | subject, pkpd1,
              labels = list(x = "Time", y = "R"),
              units = list(x = "(hrs)"))

plot(pkpd1.Gd, aspect = 1, grid = F)

f.pd1 <-
  nlme(resp ~ placebo.log(lbl, la1, la2, time)+drug.log(lemax,
  lec50, concm),
      data=pkpd1.Gd, fixed = list(lbl+la1+la2+lemax+ lec50~1),
      random = list(subject = pdDiag(list(lbl~1, lemax ~ 1))),
      start = log(c(15, .2, .1, 15, 10)),
      weight=varPower(), verbose = T)

summary(f.pd1)
exp(fixef(f.pd1))
intervals(f.pd1)

## GLMM example with binary response, Section 4.4
pkpd1$Rb <- as.integer(pkpd1$resp > 12)
mean(pkpd1$Rb)
mean(tapply(pkpd1$Rb, pkpd1$subject, mean))

f.bin1 <- glme(Rb ~ 1, pkpd1, ~1|subject, family = binomial)
summary(f.bin1)
exp(fixef(f.bin1))/(1+exp(fixef(f.bin1)))
f.bin2 <- update(f.bin1, Rb ~ concm)
summary(f.bin2)

## Simulation example of Section 4.5

# _____

# Simulate 12 trials (set pop=T)

# _____

sim <- sim.IP(dose=20, doseint=24, ndose=14,
          time=c(seq(0,1,by=0.1), seq(1,14,by=1)), nsubject=12,
          replicates=12, pop=T, seed=123)

xyplot(concm ~ time , sim, panel = panel.superpose, groups = study,
    type = "l", pch=16, lwd=1, col=c(1),layout=c(1,1),aspect=0.75,
    strip = function(...) strip.default(..., style = 1),
    par.strip.text = c(col=1), xlab = "Time (hr)", ylab = "Response")

xyplot(resp ~ time , sim, panel = panel.superpose, groups = study,
    type = "l", pch=16, lwd=1, col=c(1),layout=c(1,1),aspect=0.75,
    strip = function(...) strip.default(..., style = 1),
    par.strip.text = c(col=1),
    xlab = "Time (hr)", ylab = "Response")
```

```
# _____

# Simulate 4 studies with 12 subjects each

# _____

sim <- sim.IP(dose=20, doseint=24, ndose=14, time=c(0, 0.5, 1, 2,
3, 7, 14),
         nsubject=10, replicates=4, pop=F, seed=123)
trellis.device(graphsheet, color=F)

xyplot(concm ~ time|factor(paste("Study",study)) , sim,
      panel = panel.superpose, groups = sub, type = "b", pch=16,
      cex=0.75,
      lwd=1, lty=1, col=c(1),layout=c(2,2),aspect=0.75,
      strip = function(...) strip.default(..., style = 1),
      par.strip.text=list(cex=1.5,col=1), scales=list(cex=1.2),
      xlab = "Time (hr)", ylab = "Concentration")

xyplot(resp ~ time|factor(paste("Study",study)) , sim,
      panel = panel.superpose, groups = sub, type = "b", pch=16,
      cex=0.75,
      lwd=1, lty=1, col=c(1),layout=c(2,2),aspect=0.75,
      strip = function(...) strip.default(..., style = 1),
      par.strip.text=list(cex=1.5,col=1), scales=list(cex=1.2),
      xlab = "Time (hr)", ylab = "Response")


####################################################################
#
#
#               Modeling and simulation functions
#      Should be sourced into S-Plus before scripts are run
#
#
####################################################################

comp1.oral <-
  function(ka, ke, v, f, dose, time)
{
  (ka * dose * f)/(v * (ka - ke)) * (exp( - ke * time) - exp( -
                  ka * time))
}

comp1ss.oral <-
  function(ka, ke, v, f, dose, time, tau)
{
  (ka * dose * f)/(v * (ka - ke)) * (exp( - ke * time)/(1 - exp( - ke *
                  tau)) - exp( - ka * time)/(1 - exp( - ka * tau)))
}
```

```
comp1.oral.auc.log <-
  function(lka, lauc, lv, f, dose, time)
{
 ka <- exp(lka)
 auc <- exp(lauc)
 v <- exp(lv)
 ke <- dose/(auc * v)
 (ka * dose * f)/(v * (ka - ke)) * (exp( - ke * time) - exp( -
                  ka * time))
}
profn <-
  function(ka, ke, v, f, dose, time, tau, proftype, ndose)
{
 u0 <- (proftype == 0) * comp1.oral(ka , ke, v, f, dose, time) +
   (proftype != 0) * comp1ss.oral(ka, ke , v, f, dose, time, tau)
 if(ndose > 1) {
   nn <- ndose - 1
   for(i in 1:nn) {
     u1 <- comp1.oral(ka , ke , v, f, dose, as.double(time > i * tau) *
                   (time - i * tau))
     u0 <- u0 + u1
   }
 }
 u0
}
# _____

# Placebo model

# _____

placebo <-
  function(bl, a1, a2, time)
{
 bl + 0.5 * (1- exp(-a1*time) + exp(-a2*time) )
}
placebo.log <-
  function(lbl, la1, la2, time)
{
 bl <- exp(lbl); a1 <- exp(la1); a2 <- exp(la2)
 bl * (1- exp(-a1*time) + exp(-a2*time) )
}
# _____

# Drug model

# _____

drug <-
```

```
  function(emax, ec50, conc)
{
  (emax * conc)/(ec50+conc)
}
drug.log <-
  function(lemax, lec50, conc)
{
  emax <- exp(lemax); ec50 <- exp(lec50)
  (emax * conc)/(ec50+conc)
}
# _____

# _____

# Calculate AUC using trapezoidal rule

# _____

aucTrapEl <-
  function(data)
{
 N <- nrow(data)
 if(N == 1) return(NA)
 0.5 * sum(diff(data$time) * (data$conc[ - N] + data$conc[-1]))
}

aucTrap <-
  function(data, conc = "conc", sub = "subject", time = "time",
           dose = "dose")
{
  sub <- data[, sub]
  dataN <- data[, c(conc, time)]
  names(dataN) <- c("conc", "time")
  if(is.element(dose, names(data))) {
    sub <- paste(sub, data[, dose], sep = ":")
  }
  sub <- factor(sub, levels = unique(sub))
  val <- sapply(split(dataN, sub), aucTrapEl)
  val
}

# _____

# Extracts random effects covariance matrix from lme/nlme object

# _____

reffVar <-
  function(obj, level = 1)
{
 val <- pdMatrix(obj$modelStruct$reStruct)
 sig2 <- obj$sigma^2
```

```
  for(i in seq(along = val))
    val[[i]] <- sig2 * val[[i]]
  if(length(level) == 1) {
   val[[level]]
  }
  else {
   val[level]
  }
}


### Multilevel models
###
### Simulate data for crossover dose proportionality study
### Including possible increase in cv due to dose

sim.dp.mult <-
  function(nsub, Pars, cv.sub, cv.occ = NULL, cv.error, time, dose,
           incCV.sub = 0, incCV.occ = 0, levIncCV = 0,
           parsForm = NULL, seed = NULL)
{
  if (!is.null(seed)) set.seed(seed)
  np <- length(Pars)   # number of parameters
  nd <- length(dose)   # number of doses
  nb <- length(cv.sub) # number of subject random effects
  no <- length(cv.occ) # number of inter-occasion random effects, if any
  minD <- min(dose)
  diffD <- diff(range(dose))


  ## random effects at subject level
  reffSub <- t(diag(cv.sub) %*% array(rnorm(nsub * nb), c(nb, nsub)))
  ## random effects at occasion within-subject level
  if (no > 0) {
    reffOcc <- t(diag(cv.occ) %*% array(rnorm(nsub * no * nd),
                                        c(no, nsub * nd)))
  }
  nt <- length(time)                        # number of time points
  ## expanding random effects and covariates to match length of data
  reffSub <- reffSub[rep(1:nsub, each = nd * nt), ]
  dimnames(reffSub)[[2]] <- names(cv.sub)
  if (no > 0) {
     reffOcc <- reffOcc[rep(1:(nsub*nd), each = nt), ]
   dimnames(reffOcc)[[2]] <- names(cv.occ)
   reffOcc <- data.frame(reffOcc)
  } else {
    reffOcc <- NULL
  }
```

```
time <- rep(time, nsub * nd)
dose <- rep(rep(dose, each = nt), nsub)
sub <- rep(1:nsub, each = nd * nt)
incD <- (dose - minD)/diffD

## increasing CV of reffs, if needed
if (is.element(1, levIncCV)) {
  if (nb > 1 && (length(incCV.sub) == 1)) {
    incCV.sub <- rep(incCV.sub, nb)
  }
  ## subject level reffs
  for(i in 1:nb) {
    reffSub[,i] <- reffSub[,i] * (1 + incD * incCV.sub[i])
  }
}
reffSub <- data.frame(reffSub)
if (is.element(2, levIncCV) & (no > 0)) {
  ## occasion within-subject level reffs
  if (no > 1 && (length(incCV.occ) == 1)) {
    incCV.occ <- rep(incCV.occ, no)
  }
  for(i in 1:no) {
    reffOcc[,i] <- reffOcc[,i] * (1 + incD * incCV[i])
  }
}
## checking if any parameter is to be allowed to vary with dose
Pars <- as.vector(Pars)
ka <- Pars[1] ; ke <- Pars[2] ; v <- Pars[3]
if (!is.null(parsForm$ka)) {
 ka <- eval(parsForm$ka[[2]], list(ka = ka, dose = dose))
}
if (!is.null(parsForm$ke)) {
 ke <- eval(parsForm$ke[[2]], list(ke = ke, dose = dose))
}
if (!is.null(parsForm$v)) {
 v <- eval(parsForm$v[[2]], list(v = v, dose = dose))
}
updReff <-
  function(var, namVar, data1, data2)
{
 val <- 0
 if (!is.na(match(namVar, names(data1)))) val <- val + data1[,
 namVar]
 if (!is.null(data2) && !is.na(match(namVar, names(data2))))
   val <- val + data2[, namVar]
 var * exp(val)
}
```

```r
  ## combining fixed and random effects to form parameter values
  ka <- updReff(ka, "ka", reffSub, reffOcc)
  ke <- updReff(ke, "ke", reffSub, reffOcc)
  v <- updReff(v, "v", reffSub, reffOcc)

  ## concentrations
  conc0 <- comp1.oral(ka, ke, v, 1, dose, time)
  err <- rnorm(length(conc0), 0, cv.error)
  conc <- conc0 * (1 + err)

  val <- data.frame(subject = sub,
                    dose = dose,
                    time = time,
                    ka = ka,
                    ke = ke,
                    v = v,
                x0 = conc0,
                    conc = conc,
                    reffSub = reffSub,
                    err = err)
  if (no > 0) val$reffOcc <- reffOcc
  val
}

## multiple dose PK-PD study

sim.pkpd.mult <-
  function(nsub, doseint, ndose,
     Pars, pdPars, cv.sub, cv.occ = NULL, cv.error, time, dose,
        incCV.sub = 0, incCV.occ = 0, levIncCV = 0,
        parsForm = NULL, seed = NULL)
{
  if (!is.null(seed)) set.seed(seed)
  np <- length(Pars)    # number of parameters
  nd <- length(dose)    # number of doses
  nb <- length(cv.sub)  # number of subject random effects
  no <- length(cv.occ)  # number of inter-occasion random effects,
                           if any
  minD <- min(dose)
  diffD <- diff(range(dose))
## random effects at subject level
reffSub <- t(diag(cv.sub) %*% array(rnorm(nsub * nb), c(nb, nsub)))
## random effects at occasion within-subject level
if (no > 0) {
  reffOcc <- t(diag(cv.occ) %*% array(rnorm(nsub * no * nd), c(no,
  nsub * nd)))
}
nt <- length(time)      # number of time points
```

```
## expanding random effects and covariates to match length of data
reffSub <- reffSub[rep(1:nsub, each = nd * nt), ]
dimnames(reffSub)[[2]] <- names(cv.sub)
if (no > 0) {
   reffOcc <- reffOcc[rep(1:(nsub*nd), each = nt), ]
  dimnames(reffOcc)[[2]] <- names(cv.occ)
  reffOcc <- data.frame(reffOcc)
} else {
   reffOcc <- NULL
}

time <- rep(time, nsub * nd)
dose <- rep(rep(dose, each = nt), nsub)
sub <- rep(1:nsub, each = nd * nt)
incD <- (dose - minD)/diffD

## increasing CV of reffs, if needed
if (is.element(1, levIncCV)) {
  if (nb > 1 && (length(incCV.sub) == 1)) {
    incCV.sub <- rep(incCV.sub, nb)
  }
  ## subject level reffs
  for(i in 1:nb) {
    reffSub[,i] <- reffSub[,i] * (1 + incD * incCV.sub[i])
  }
}
reffSub <- data.frame(reffSub)
if (is.element(2, levIncCV) & (no > 0)) {
  ## occasion within-subject level reffs
  if (no > 1 && (length(incCV.occ) == 1)) {
    incCV.occ <- rep(incCV.occ, no)
  }
  for(i in 1:no) {
    reffOcc[,i] <- reffOcc[,i] * (1 + incD * incCV[i])
  }
}

## checking if any parameter is to be allowed to vary with dose
Pars <- as.vector(Pars)
ka <- Pars[1] ; ke <- Pars[2] ; v <- Pars[3]
bl <- Pars[4] ; ec50 <- Pars[5] ; emax <- Pars[6]
pdPars <- as.vector(pdPars)
a1 <- pdPars[1]; a2 <- pdPars[2]

if (!is.null(parsForm$ka)) {
  ka <- eval(parsForm$ka[[2]], list(ka = ka, dose = dose))
}
if (!is.null(parsForm$ke)) {
  ke <- eval(parsForm$ke[[2]], list(ke = ke, dose = dose))
```

```
}
if (!is.null(parsForm$v)) {
  v <- eval(parsForm$v[[2]], list(v = v, dose = dose))
}
if (!is.null(parsForm$bl)) {
  bl <- eval(parsForm$bl[[2]], list(bl = bl, dose = dose))
}
if (!is.null(parsForm$ec50)) {
  ec50 <- eval(parsForm$ec50[[2]], list(ec50 = ec50, dose = dose))
}
if (!is.null(parsForm$emax)) {
  emax <- eval(parsForm$emax[[2]], list(emax = emax, dose = dose))
}
updReff <-
    function(var, namVar, data1, data2)
{
    val <- 0
    if (!is.na(match(namVar, names(data1)))) val <- val + data1[,
     namVar]
    if (!is.null(data2) && !is.na(match(namVar, names(data2))))
     val <- val + data2[, namVar]
    var * exp(val)
}
## combining fixed and random effects to form parameter values
ka <- updReff(ka, "ka", reffSub, reffOcc)
ke <- updReff(ke, "ke", reffSub, reffOcc)
v <- updReff(v, "v", reffSub, reffOcc)

## combining fixed and random effects to form parameter values for PD model
bl <- updReff(bl, "bl", reffSub, reffOcc)
ec50 <- updReff(ec50, "ec50", reffSub, reffOcc)
emax <- updReff(emax, "emax", reffSub, reffOcc)

## concentrations
conc0 <- comp1.oral(ka, ke, v, 1, dose, time)
err <- rnorm(length(conc0), 0, cv.error)
conc <- conc0 * (1 + err)
concm <-    profn(ka,ke,v,1,dose,time,doseint,proftype=0,ndose)
placebo <- placebo(bl, a1, a2, time)
drug <- drug(emax, ec50, concm)
err <- rnorm(length(concm), 0, cv.error)
resp <- (placebo + drug) * (1 + err)

val <- data.frame(subject = sub,
                  dose = dose,
                  time = time,
                  ka = ka,
                  ke = ke,
```

```
                     v = v,
                     bl=bl,
                     ec50=ec50,
                     emax=emax,
                     a1=a1,
                     a2=a2,
                     x0 = conc0,
                     conc = conc,
                     concm=concm,
                     placebo=placebo,
                     drug=drug,
                     resp=resp,
                     reffSub = reffSub,
                     err = err)
  if (no > 0) val$reffOcc <- reffOcc
  val
}
# _____

# Fixed effects and corresponding covariance matrix

# _____

getFixPars <-
  function(object)
{
  ## fixed effects estimates and var-cov matrix
  list(coef = fixef(object), var = object$varFix)
}
# _____

# Variance-covariance  components  and  corresponding  covariance
  matrix

# _____

getRanPars <-
  function(object)
{
  ## variance-covariance components estimates and var-cov matrix
  aux <- object$apVar
  if (!is.numeric(aux)) stop(aux)
  val <- list(coef = attr(aux, "Pars"))
  attr(aux, "Pars") <- attr(aux, "natural") <- attr(aux, "natUn-
  cons") <- NULL
  val$var <- aux
  val
}
### simulate parameters according to estimated distribution from
    NLME
```

```
simPars <-
  function(object, N = 1)
{
  ## simulate parameters according to mean = coef, var = var
  val <- rmvnorm(N, mean = object$coef, cov = object$var)
  dimnames(val) <- list(1:N, names(object$coef))
  val
}


convRanPars <-
  function(pars)
{
  ## converts unconstrained simulated values for var-cov
  ## components into "natural" parameters, in DP example
  N <- length(pars)
  sig <- exp(pars[N])
  power <- pars[N - 1]
  pars <- pars[1:(N-2)]
  pars <- exp(pars)
  pars[c(4, 8)] <- (pars[c(4,8)] - 1)/(pars[c(4,8)] + 1)
  names(pars) <- rep(c(«sd(lKa)», «sd(lAUC)», «sd(lV)», «cor(lAUC, lV)»),2)
  list(Subject = pars[5:8],
      "Dose %IN% Subject" = pars[1:4],
      power = power,
      sigma = sig)
}


# _____


# Simulate clinical response in a trial.
# nsub represents number of subjects
# replicates represents the number of clinical studies to be simulated.
# The parameter pop if set = T will calculate population estimates.
# In this case subject random effects are set to zero, and
# number of subjects is forced = 1.
# _____


sim.IP <-
  function(dose, doseint, ndose, time, nsubject, replicates, pop, seed=NULL)
{
  nsub = nsubject
  if(pop) {nsub <- 1}
  dose <- dose # mg
  doseint <- doseint # hours
  ndose   <- ndose
  ## Invoke fixed and random effects for the PK and PD models
```

```
pkobj   <- f.dp1b
fxPk <- getFixPars(pkobj)
rnPk <- getRanPars(pkobj)


pdobj <- f.pd1
fxPd <- getFixPars(pdobj)
rnPd <- getRanPars(pdobj)


## _____
if(!is.null(seed)) set.seed(seed)
sub <- 1:nsub
nstudy <- replicates
study <- 1:nstudy
time0      <- time       # days
time0      <- time0 * 24 # convert to hours


## _____


## Generate concentrations at SS using pk estimates of object
   f.dp1.d
simfxPk <- simPars(fxPk, N = nstudy)  # samples from pd model,
                                      # i.e. bl a1, a2, emax and ec50
simrnPk <- simPars(rnPk, N = nstudy)  # samples from random effects
                                      #(in bl, emax and ec50),
                                      # power parameter and sigma


# Generate response variable using pkpd estimates of object f.pd1.b
simfxPd <- simPars(fxPd, N = nstudy)   # samples from pd model,
                                       # i.e. bl a1, a2, emax and ec50
simrnPd <- simPars(rnPd, N = nstudy)   # samples from random effects
                                       #(in bl, emax and ec50),
                                       # power parameter and sigma


## _____
##   lauc <- int + beta * log(dose)
##   ke <- dose/(auc * v)
lauc <- simfxPk[,3] + simfxPk[,4] * log(dose)
lka <- simfxPk[,1]
ke <- dose/(exp(lauc) * exp(simfxPk[,2]))
lv <- simfxPk[,3]
lbl <- simfxPd[,1]                         # bl
la1 <- simfxPd[,2]                         # a1
la2 <- simfxPd[,3]                         # a2
lemax      <- simfxPd[,4]                      # emax
lec50      <- simfxPd[,5]                      # ec50
```

```
lauc <- t(array(lauc, c(nstudy, nsub)))
lka <- t(array(lka , c(nstudy, nsub)))
lv <- t(array(lv , c(nstudy, nsub)))
lbl <- t(array(lbl , c(nstudy, nsub)))
la1 <- t(array(la1 , c(nstudy, nsub)))
la2 <- t(array(la2 , c(nstudy, nsub)))
lemax <- t(array(lemax, c(nstudy, nsub)))
lec50 <- t(array(lec50, c(nstudy, nsub)))

if(!pop){
  cv.ka <- exp(simrnPk[,1])
  cv.auc <- exp(simrnPk[,2])
  cv.v <- exp(simrnPk[,3])
  cv.bl <- exp(simrnPd[,1])
  cv.emax <- exp(simrnPd[,2])
  cv.ec50 <- exp(simrnPd[,3])

  if(nstudy==1) {
    ran.ka <- t(cv.ka %*% array(rnorm(nsub * nstudy), c(nstudy,
                                  nsub)))
    ran.auc <- t(cv.auc %*% array(rnorm(nsub * nstudy), c(nstudy,
                                   nsub)))
    ran.v <- t(cv.v %*% array(rnorm(nsub  *  nstudy),  c(nstudy,
                                 nsub)))
    ran.bl <- t(cv.bl %*% array(rnorm(nsub * nstudy), c(nstudy,
                                  nsub)))
    ran.emax <- t(cv.emax %*% array(rnorm(nsub*nstudy),c(nstudy,
                                     nsub)))
    ran.ec50 <- t(cv.ec50 %*% array(rnorm(nsub*nstudy),c(nstudy,
                                     nsub)))
}

  if(nstudy>1) {
    ran.ka <- t(diag(cv.ka) %*% array(rnorm(nsub    *    nstudy),
                                  c(nstudy, nsub)))
    ran.auc <- t(diag(cv.auc) %*% array(rnorm(nsub   *   nstudy),
                                    c(nstudy, nsub)))
    ran.v <- t(diag(cv.v) %*% array(rnorm(nsub*nstudy),c(nstudy,
                                   nsub)))
    ran.bl <- t(diag(cv.bl) %*% array(rnorm(nsub    *    nstudy),
                                  c(nstudy, nsub)))
    ran.emax<- t(diag(cv.emax) %*% array(rnorm(nsub   *   nstudy),
                                     c(nstudy, nsub)))
    ran.ec50 <- t(diag(cv.ec50) %*% array(rnorm(nsub  *  nstudy),
                                      c(nstudy, nsub)))
  }

  lka <- lka+ran.ka
  lauc <- lauc+ran.auc
  lv <- lv+ran.v
```

```
  lbl <- lbl+ran.bl
  lemax <- lemax+ran.emax
  lec50 <- lec50+ran.ec50
}

lke <- log(dose) - (lauc + lv)

d1 <- length(time0)
d2 <- nsub
d3 <- nstudy
concm <- array(double(d1*d2*d3),dim=c(d1,d2,d3))
resp <- array(double(d1*d2*d3),dim=c(d1,d2,d3))

## compute Response profiles
for (n in 1:nstudy) {
  for (m in 1:nsub) {
    for (i in 1:length(time0)) {
      ## Conc entration profile
      concm[i,m,n] <-
       profn(exp(lka[m,n]),exp(lke[m,n]),exp(lv[m,n]),1,dose,time
       0[i], doseint,proftype=0,ndose)
      ## Response profile
      resp[i,m,n] <-
       placebo.log(lbl[m,n], la1[m,n], la2[m,n], time0[i]) +
         drug.log(lemax[m,n], lec50[m,n], concm[i,m,n])
    }
  }
}
data.frame(time=rep(time0,nsub*nstudy),
           study=rep(study,each=length(time0)*nsub),
           sub=rep(rep(sub, each=length(time0)),nstudy),
           concm =as.vector(concm),
           resp=as.vector(resp))
}
```

Bayesian Hierarchical Modeling with Markov Chain Monte Carlo Methods

STEPHEN B. DUFFULL, LENA E. FRIBERG, and CHANTARATSAMON DANSIRIKUL

5.1 INTRODUCTION

5.1.1 Background

There is a large volume of literature that deals with Bayesian ideas and methods of data analysis, decision analysis, and design. Much of this literature is highly technical and arises from specialized settings; for example, Gibbs sampling, which is now recognized as an important Bayesian tool, originally arose based on solving a problem in engineering in the early 1980s (1). In this chapter we focus on practical applications of Bayesian methods for analysis of data that arise from either pharmacokinetic (PK) or pharmacokinetics/pharmacodynamic (PK/PD) studies. Where examples of model-based notation and code are provided, we have done so based on the general structure used by WinBUGS. This chapter is divided into four main sections. The first section provides a brief introduction to Bayesian hierarchical modeling. In Section 5.2, we provide a *how to* for defining priors. In Section 5.3, we introduce methods for model discrimination in a Bayesian setting. Finally, in Section 5.4, we provide a summary.

Before embarking on our initial aim of describing Bayesian hierarchical modeling in a pharmacometric setting, it is worth devoting a few words to the notion of *what constitutes a Bayesian analysis*. At a first glance it may seem obvious as to what constitutes a Bayesian analysis (simply anything called "Bayesian" would be a good initial point for categorizing methods). Unfortunately, this simplification is not always the case; nor is it particularly helpful. Compare, for instance, empirical Bayes's estimates of parameters versus Bayesian estimates (e.g., see the POSTHOC option relating to the "FO" method in NONMEM (2)). In accordance with Bayes's expression (Eq. (5.1)), the Bayesian approach involves the incorporation of prior beliefs about the parameters θ, given by $\pi(\theta)$, with study outcomes (Y). The study outcomes are expressed as the likelihood of the data given the model and

parameter values $\pi(Y \mid \theta)$ (termed the likelihood), to provide updated beliefs about the parameters $\pi(\theta \mid Y)$, termed posterior beliefs.

$$\pi(\theta|Y) = \frac{\pi(Y|\theta)\pi(\theta)}{\pi(Y)} \tag{5.1}$$

We have used the greek letter π as it is not used to represent its numerical value in this chapter—and (in an abuse of notation) in the above expression it can be used to mean either the *probability of* (elsewhere we have used Pr) or the *probability distribution of* (elsewhere we have used $p(\ )$). The denominator, $\pi(Y)$, is the marginal distribution of the data and is in essence unable to be quantified for PK or PK/PD analyses (it is equal to the multiple integral of the numerator over all parameters). Hence, Bayes's expression is often written to the level of proportionality $\pi(\theta \mid Y) \propto \pi(Y \mid \theta)\pi(\theta)$. When interpreting Bayes's expression, it appears that the fundamental principle of the Bayesian approach is to learn about some experiment as the "weighted average" of some prior beliefs and observations that arise from the actual experiment itself. However, if the prior were set to be essentially noninformative (i.e., uniform over the plausible range of parameter values), then the posterior is proportional to the likelihood, that is, $\pi(\theta \mid Y) \propto \pi(Y \mid \theta)$, and no influence of the prior will be discernable. Yet the analysis could still be Bayesian. The corollary of this consideration is where priors are used but the analysis is termed non-Bayesian (e.g., see Ref. 3). In this particular case, the investigators used prior information formally in their analysis but claimed the analysis to be non-Bayesian. So, it can be concluded that including a prior is not exclusive to Bayesian analyses. One attribute that is nonfrequentist is the consideration that parameter values themselves are random variables that arise from some unknown distribution, which contrasts with nonBayesian theory where it is believed that there is only a single true set of parameter values that solve for the data and model. This then leads to confusion over whether the maximum a posteriori method of Bayesian forecasting (see Ref. 4) is truly Bayesian since the goal is to locate a point estimate at the mode of the posterior distribution of the parameters rather than the full distribution. Some may consider this approach to be empirical Bayes.

It is not our goal, however, to present all sides to this argument but rather to make the reader aware that many contradictions exist and that application of various methods is perhaps a sufficient goal. For the purposes of this chapter, we refer to *fully* Bayesian methods as those that give rise to knowledge about the full posterior distribution of the parameters, which is a function of the joint distribution of the likelihood and the prior. This process allows for the uncertainty in the parameters to be modeled explicitly.

5.1.2 Bayesian Methods for Population PK/PD Analysis: Hierarchical Modeling

Similar to the non-Bayesian framework for analysis of repeated measures data, the Bayesian setting also shares the same format for describing stages 1 and 2 of the hierarchical model but has the addition of the third stage assigned to specification of the priors (see Ref. 5 for an in-depth discussion of the hierarchical framework for analysis, and for a comparison between MCMC and maximum likelihood methods

readers are referred to Ref. 6). In theory, the Bayesian approach can accommodate any number of levels in the hierarchy that may represent different levels associated with the random effects terms. An obvious example is the inclusion of between-occasion variability (see Lunn and Aarons (7) for a description) to make a four-stage hierarchical model. For simplicity, a standard three-stage hierarchical model is shown. The choice of probability density functions for each of the stages is discussed in more detail in Section 5.2, where specification of the prior is considered.

Stage 1—Model for the Data

$$y_{ij}^{iid} \sim N(f(\boldsymbol{\theta}_i, x_{ij}), \sigma^2) \tag{5.2}$$

Here the data y_{ij}, the jth observation for the ith subject, are assumed to be known and independently normally distributed around the model prediction $f(\boldsymbol{\theta}_i, x_{ij})$ with variance σ^2. $\boldsymbol{\theta}_i$ represents a vector of individual parameter values for the ith individual and x_{ij} is the sampling time.

Stage 2—Model for Heterogeneity Between Subjects

$$\boldsymbol{\theta}_i \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Omega}) \tag{5.3}$$

Here $\boldsymbol{\theta}$ is a vector of mean population pharmacokinetic parameters and $\boldsymbol{\Omega}$ is the variance–covariance matrix of between-subject random variability. N_p represents a p-dimensional multivariate normal distribution, where p is the number of parameters. It is often more useful to consider the values of the parameters for the individual to be related to the population parameters via a covariate relationship, in which case the expression may be written as

$$\boldsymbol{\theta}_i \sim N_p(g(\boldsymbol{\theta}, \mathbf{z}_i), \boldsymbol{\Omega}) \tag{5.4}$$

In this notation, $g(\boldsymbol{\theta}, \mathbf{z}_i)$ is used to represent a function (g), perhaps a linear combination of covariates, that describes the expectation of the ith subjects parameter vector $\boldsymbol{\theta}_i$ conditional on their demographic characteristics $(\mathbf{z}_i)$ and population parameter values $(\boldsymbol{\theta})$. The variance-covariance matrix $(\boldsymbol{\Omega})$ therefore describes the random variability between subjects that is not able to be explained by covariates.

Stage 3—Model for the Priors   The third stage involves specification of the prior structure (this is discussed in more detail in Section 5.2). Typically, the prior for a model for PK and PD parameters would be set up for the residual uncertainty as

$$\sigma \sim U(a, b) \tag{5.5}$$

for the mean parameter values,

$$\boldsymbol{\theta} \sim N_q(\bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) \tag{5.6}$$

and for the precision of heterogeneity,

$$\boldsymbol{\Omega}^{-1} \sim Wi_p(\rho \boldsymbol{\Omega}_0, \rho), \quad \text{and} \quad \rho \geq p \text{ and } p \leq q \tag{5.7}$$

The prior for the residual standard deviation is given here by a uniform distribution over the range a to b, where a is typically very small (e.g., 0.0001) and b is sufficiently large as to encompass extreme values of the response variable. The mean parameter values, of which there are q, are shown to be distributed according to a multivariate normal distribution with a hyperprior mean $\bar{\mu}$ and variance–covariance matrix Σ that describes the uncertainty with which we know θ. It should be remembered that Σ is *not* the uncertainty of $\bar{\mu}$. The prior for the inverse of the variance–covariance matrix Ω^{-1} describing between-subject heterogeneity is given by a p-dimensional Wishart distribution with parameters Ω_0 and ρ. Ω_0 is the estimate of the prior expectation of the variance–covariance matrix and ρ is the degrees of freedom of the Wishart distribution. The minimum allowable value of ρ is p (which is least informative), and higher values can be chosen depending on the level of informativeness that is desired. This notation allows for there to be some population parameters that do not have variability between subjects (i.e., for $q > p$).

Specification of the values of a, b, $\bar{\mu}$, Σ, Ω_0, and ρ is at the discretion of the pharmacometrician (see Section 5.2 for details). The choice of a different prior structure, for example, using a multivariate-t distribution rather than normal, may also be appropriate if there is some evidence suggesting the presence of potential outlying subjects.

It is worth mentioning at this stage that the three-stage hierarchical model used in Bayesian analyses when undertaken within the framework provided by WinBUGS requires that normal distributions are parameterized as mean and precision. Precision is the inverse of variance. For example, when defining the prior for the population parameter vector θ, the multivariate normal distribution would be parameterized as the mean vector $\bar{\mu}$ and the inverse of the variance–covariance matrix Σ^{-1} such that,

$$\theta \sim N_q\,(\bar{\mu}, \Sigma^{-1})$$

This parameterization would also hold for Eq. (5.3), (5.4), and (5.6). Hereafter the notation of WinBUGS is adopted and where possible examples of WinBUGS code are provided. Readers are referred to Fryback et al. (8) for an introduction to WinBUGS (see especially the appendixes for reference to how to run WinBUGS), Duffull et al. (6) for an introductory population PK example, and Lunn et al. (9) for a more in-depth treatment.

5.1.3   Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are a group of methods that can be used to explore the posterior distribution of the parameters (e.g., of a PK or PK/PD model) conditional on some observable quantities. They can also be used in non-Bayesian settings. There are two main MCMC techniques used (readers are referred to Gilks et al. (10) for an in-depth overview of MCMC techniques), namely, the Metropolis–Hastings (MH) algorithm, originating from Metropolis et al. (11) and generalized by Hastings (12) and Gibbs sampling (1). Indeed, Gibbs sampling and many other MCMC methods are special cases of the MH algorithm.

MCMC methods are essentially Monte Carlo numerical integration that is wrapped around a purpose built Markov chain. Both Markov chains and Monte Carlo integration may exist without reference to the other. A Markov chain is any chain where the current state of the chain is conditional on the immediate past state only—this is a so-called first-order Markov chain; higher order chains are also possible. The chain refers to a sequence of realizations from a stochastic process. The nature of the Markov process is illustrated in the description of the MH algorithm (see Section 5.1.3.1).

Monte Carlo integration is a process characterized by the use of random sampling often for integration. The premise underpinning the idea is remarkably simple, in that if it is possible to generate sufficient numbers of random samples (e.g., L samples) from a distribution (that may have an unknown form), then the underlying distribution can be explored by pooling those samples and the mean calculated accordingly. For example, if X is used to denote a random deviate from $f(X)$, then

$$E[X] \approx L^{-1} \sum_{l=1}^{L} f(X_l) \tag{5.8}$$

which as $L \to \infty$,

$$E[X]_{\lim L \to \infty} = L^{-1} \sum_{l=1}^{L} f(X_l) = \int f(X)\, dX \tag{5.9}$$

The important conceptual point is that the parametric form of $f(X)$ does not need to be known for either the expectation of X to be calculated or in order that random deviates may be generated from $f(X)$. The only requirement is that the value of $f(X)$ must be evaluable at all (legal) values of X; it is not a requirement that the integral of $f(X)$ (as in Eq. (9)) be able to be computed in closed form. It is also not required that $f(X)$ be a univariate distribution, but $f(\mathbf{X})$ may represent a joint distribution of $\mathbf{X}$. The marginal distribution of $f(X)$ may be "extracted" from the joint distribution using exactly the same random sampling technique but keeping each marginal set of random deviates separately. A joint distribution is a combined distribution of many parameters and the parameters may or may not share the same distributional form. One simple method of generating random samples from this distribution is that of rejection sampling (see Press et al. (13) for an overview of random sampling, Smith and Gelfand (14) for an introduction to rejection sampling for Bayesian analyses, and Wakefield (15) for an application). The MH method and Gibbs sampling also provide methods for generating random deviates from $f(X)$.

5.1.3.1   *Metropolis–Hastings Algorithm*

The Metropolis–Hastings algorithm is the most general form of the MCMC processes. It is also the easiest to conceptualize and code. An example of pseudocode is given in the five-step process below. The Markov chain process is clearly shown in the code, where samples that are generated from the prior distribution are accepted as arising from the posterior distribution at the ratio of the probability of the joint

distribution evaluated at the current set of parameters over the probability evaluated at the last best (but not "best ever") set of parameters (shown in step 3). The joint distribution is provided by the product of the prior probability of the random set of parameters with the likelihood.

Step 1.	Set	$i = 1$	# counter
Step 2.	While $1 < i <$ max		
	do Steps 3–5		
Step 3.	Sample	$u \sim U(0, 1)$	# acceptance probability
		$\boldsymbol{\beta}^i \sim p(\boldsymbol{\beta})$	# sample parameters from prior
	Set	$n = p(\mathbf{y}\|\boldsymbol{\beta}^i)$	# joint prior-likelihood
		$d = p(\mathbf{y}\|\boldsymbol{\beta}^{i-1})$	
		$a_1 = \Pr(\boldsymbol{\beta}^{i-1})$	
		$a_2 = \Pr(\boldsymbol{\beta}^i)$	
		$ratio = n/d \times a_1/a_2$	# acceptance ratio
Step 4.	if $ratio > u$	$\boldsymbol{\beta}^{\text{last}} = \boldsymbol{\theta}^i = \boldsymbol{\beta}^i$	# accept samples
			# reset new best value
	Else set	$\boldsymbol{\theta}^i = \boldsymbol{\beta}^{\text{last}}$	# keep previous best samples
Step 5.	$i = i + 1$		# increment counter
Step 6.	OUTPUT($\boldsymbol{\theta}$)		
	STOP.		

5.1.3.2 *Gibbs Sampling Algorithm*

Gibbs sampling is a specialized and more efficient version of the MH algorithm. In this procedure there are no rejected samples, the sampling distributions are set up so that once the chain has settled down (to the so-called stationary distribution) all samples are considered to arise from the posterior distribution. The cost of this improvement in the process is that the user needs to define the conditional sampling distribution, which includes distributions for all remaining parameters and the data, in closed form (see step 3). This is analytically possible for linear models and for combinations of the prior and likelihood that are conjugate, meaning that the posterior distribution will have the same structural form as the prior but with updated parameter values (see Fryback et al. (8) for a brief and elementary explanation of this process). For example, for consideration of a simple univariate distribution if the prior distribution of the parameters is normal and the likelihood is normal, the posterior distribution will also be normal. In circumstances where nonconjugate priors are chosen—that is, the prior and likelihood do not arise from the same and conjugate family of distributions or the model is nonlinear in its parameters—then an MH step may be required to be performed within the overall Gibbs sampling process and hence hybrid MCMC procedures arise. Gibbs sampling is shown below in pseudocode for a three-parameter model.

Step 1.	Set	$\beta^1 = \beta^0$	# set the parameter values = initial estimates
		$i = 1$	# initialize counter
Step 2.	While $i <$ max do Step 3		
Step 3.	Sample	$\beta^i_1 \sim p(\beta^i_1 \mid \beta^{i-1}_2, \beta^{i-1}_3, \mathbf{y})$	# sample parameter 1
		$\beta^i_2 \sim p(\beta^i_2 \mid \beta^i_1, \beta^{i-1}_3, \mathbf{y})$	# sample parameter 2
		$\beta^i_3 \sim p(\beta^i_3 \mid \beta^i_1, \beta^i_2, \mathbf{y})$	# sample parameter 3
	Set	$i = i + 1$	# increment counter
Step 4.	OUTPUT (β) STOP		

5.1.3.3 Diagnostics

Model Diagnostics   Graphical model assessments such as predictions versus observations, weighted residuals versus time, and weighted residuals versus predictions are, as in non-Bayesian analyses, valuable for model development. In WinBUGS the user needs to provide the code for population predictions and weighted residuals. Correlation plots can be generated directly in WinBUGS Version 1.4 without the need to export the data into another program. However, the CODA function can create a file with all sampled values from each chain of each parameter, which can be exported to other programs (e.g., S-Plus, R, Matlab) for diagnostic purposes.

Convergence Diagnostics   Inferences from the posterior distributions should be made after convergence has been achieved to assure that the posterior distributions represent the target distributions. However, there is no diagnostic method that can be used to provide a guarantee that convergence has occurred. Therefore, it is recommended that several methods are used. One method is to visualize the histories of the chains against the iteration number (see Duffull et al. (6) and Lunn et al. (9) for visual examples). Such plots should look like "fuzzy caterpillars." If the appearance of the history is a "wiggly snake," it (generally) indicates that the sampler needs to be run longer and/or that the model needs to be reparameterized. The so-called wiggly snake is associated with serial correlation in the sampling chain (termed autocorrelation). It has been recommended to run at least two chains simultaneously, with overdispersed initial estimates (e.g., let the initial estimates of chain 2 be 50% higher than the initial estimates of chain 1). If the histories of the chains are overlapping and appear to mix with each other, then this is an indication of convergence, but does not assure convergence. Although no direct statistics can be applied to visualization of the chain, it is intuitively appealing in its simplicity and does from experience identify many chain convergence issues.

A more objective method is to investigate the Gelman–Rubin diagnostics for chain convergence. This procedure is automated within WinBUGS. This method compares the between-chains and within-chain variability in a similar spirit to an analysis of variance. Samples are required from at least two chains that are started

with overdispersed initial values. The criterion for convergence is achieved when the ratio between the 80% interval of the pooled chains and the 80% interval within the chains (averaged over all chains) is close to 1 for all parameters of interest. Again, no specific criteria are available that show a definite convergence or lack of convergence. Another diagnostic is Geweke's method. This can easily be computed from an output analysis of the chains; for example, the Bayesian Output Analysis (BOA) program has this diagnostic as an automated feature, but it is not available automatically from within WinBUGS. This method compares the sample distributions in the first half of the samples within a chain with the sample distribution from the samples from the second half of the chain. Logically these sample distributions should be indistinguishable.

Finally, the Monte Carlo error (MC error) can be used to assess how many iterations need to be run after convergence for accurate inference from the posterior distribution. The MC error is an estimate of the deviance between the mean of the sampled values and the posterior mean; this error can be likened to a standard error. Generally, an MC error of less than 5% of the sample standard deviation of the parameters of interest is recommended.

Autocorrelation   Ideally, all samples from a chain should be independent, that is, free from serial correlation. However, in reality, this is rarely the case. The presence of autocorrelation does not indicate either a lack of convergence or necessarily overparameterization (although reparameterization or a reduction in the dimensionality of the model will often reduce or eliminate autocorrelation). It will be necessary, however, to run the chains for longer so that ultimately enough "independent" samples from the chain are kept to ensure that the posterior distribution has been suitably explored by the sampler. The influence of autocorrelated samples in the posterior distribution may be reduced by thinning, where only a fraction of the samples from the posterior distribution are kept; for example, it is common to retain only every tenth sample. As a result ten times more samples are needed to generate the same number of samples from the posterior distribution. Thinning does not change the occurrence of autocorrelation, but it does reduce the apparent influence of autocorrelation since setting thinning to 10 results in 90% of the samples being discarded. Thinning is used as a method of saving computer memory by allowing chains to be run longer without the need to save every sample. This is often needed for analysis of large data sets, as may occur in PK/PD analyses. It is of course preferable not to thin samples unless absolutely required.

5.2 SPECIFICATION OF PRIORS

5.2.1 Defining Bayesian Priors

As mentioned earlier, incorporating prior information does not in itself constitute a Bayesian approach. Priors have been used in non-Bayesian settings in population PK analysis and other analyses. Applications using the PRIOR subroutine in NONMEM have been described previously (3, 16). In this setting the prior information can be viewed as a penalty on the likelihood function, and its implementation is similar in spirit to the maximum a posteriori (MAP) procedures used commonly

in Bayesian forecasting programs. Approaches based on the "frequentist-prior," however, are relatively uncommon. In contrast, Bayesian analyses always include prior information as it is explicitly defined in Bayes's expression, although the priors can be set to be in principle noninformative.

Defining the values of the priors and their informativeness is therefore an essential part of any Bayesian analysis.

In this chapter we consider that specification of priors may be divided into three broad and useful categories. The first provides a general overview of the setup of priors for a Bayesian PK (or PK/PD) analysis, while outlining the so-called noninformative (sometimes referred to as vague) priors, which represent the application of priors that are not intended to influence the analysis and arise from the belief that we know almost nothing about the manner in which the current data may have arisen. The second category refers to what we have termed biologically plausible but low-information priors. These are priors that have fairly vague information but are limited to span an interval that would seem biologically plausible for most PK/PD analyses. The third category refers to the development of informative priors.

Note that, whatever the final model parameterization used for the current analysis, it must be constructed in such a manner as to forcibly eliminate models and model parameterizations that contain inherent identifiability problems. This is relevant to all prior structures. The most common source in pharmacokinetics is the so-called flip-flop models, where the model predictions are identical for two or more sets of solutions of the parameters; for example, where K_a and CL/V_d are exchangeable within the model, such that any set of values of K_a may be exchanged with the ratio CL/V_d to provide the same response values. Model parameterization that contains flip-flop characteristics can result in poor chain mixing during the MCMC process, such that the chain flips between one or another solution, potentially inducing artificial bimodal posterior distributions. PKBUGS performs this reparameterization automatically, thereby forcing K_a to be greater than CL/V_d—although this may not always be desirable.

5.2.2   Noninformative Priors

Many Bayesian analyses utilize so-called noninformative priors (see examples in the WinBUGS manual (17)). The principal belief underlying their wide utility is to retain objectivity in relation to the current analysis. Should prior evidence influence the analysis of the current experiment, then the objectivity of the current analysis may be questioned, due to the subjective nature of priors and methods for their elicitation. In a philosophical sense, it might also be argued that it is equally nonobjective to ignore all previous evidence, no matter how applicable or strong the evidence might be.

The use of noninformative priors itself is not without its difficulties, by the simple virtue that truly noninformative proper priors do not really exist. However, with very low precision terms (e.g., 0.0001) and an assumption of lognormality appropriate for many PK/PD parameters, the priors can be considered very vague.

In defining priors there are two main considerations: first, the choice of the prior distribution and second, the choice of its parameter values (we only consider parametric priors). The choice of the prior distribution itself is not trivial as it is

essential that the random deviates generated from this distribution will have the right characteristics; for example, if you are generating random deviates to be considered as candidates of a variance parameter, then the deviates must be positive numbers. In addition to choosing a prior that has the right distributional properties, it is convenient to also consider priors that will ease the computational burden during the MCMC process. For instance, choosing conjugate priors (i.e., those that have distributional characteristics that can be combined algebraically with the specified distribution of the likelihood) greatly facilitates the speed of the MCMC process. Fortunately, for PK and PK/PD analyses, there are a standard set of conjugate distributions that are commonly used for generating random candidates of parameter values and it is generally just a matter of choosing the parameter values for each distribution.

In the following notation we assume that the population mean values of a vector of parameters (θ) are distributed multivariate normally with some uncertainty depicted as a variance–covariance matrix (Σ). It is convention in WinBUGS to express variability as precision (the inverse of variance); hence, the precision of θ is given by Σ^{-1}. This prior can be cast as a lognormal distribution by exponentiating the individual values of the parameters at stage 1 of the hierarchical model. Some authors have chosen to use a t-distribution with small degrees of freedom (to allow for heavy tails) instead of a normal prior. While this may be more flexible if outlying subjects are suspected, the t-distribution is not conjugate with the normal likelihood, which greatly adds to the computation time. For a model with three fixed effects parameters, the following vague prior structure can be used:

$$\boldsymbol{\theta} \sim N(\bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}^{-1}), \quad \bar{\boldsymbol{\mu}} = \begin{bmatrix} \bar{\mu}_{\theta 1} \\ \bar{\mu}_{\theta 2} \\ \bar{\mu}_{\theta 3} \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.0001 & 0 & 0 \\ 0 & 0.0001 & 0 \\ 0 & 0 & 0.0001 \end{bmatrix} \tag{5.10}$$

Note that the parameter vector $\bar{\mu}$ is usually expressed as the natural logarithm of the parameter values. The prior for the variance–covariance matrix of the between-subject effects ($\boldsymbol{\Omega}$), is usually given by a Wishart distribution (which is conjugate with the normal distribution). Simulation from the Wishart distribution will produce inverse candidates of $\boldsymbol{\Omega}$, which can be thought of as a precision matrix of between-subject heterogeneity. The Wishart distribution is parameterized in terms of ρ (degrees of freedom) and the estimate of the mean of $\boldsymbol{\Omega}$ (denoted by $\boldsymbol{\Omega}_0$, where 0 is used to signify that this is an initial estimate). The informativeness of the Wishart—that is, how similar the simulated values of $\boldsymbol{\Omega}$ are to $\boldsymbol{\Omega}_0$—depends on the number of degrees of freedom, where the number equal to the size of the matrix is least informative (i.e., for a 3×3 matrix $\rho = 3$ is least informative).

$$\boldsymbol{\Omega}^{-1} \sim Wi(\rho\boldsymbol{\Omega}_0, \rho), \quad \rho = 3 \tag{5.11}$$

The Wishart distribution is a multivariate gamma distribution, which itself is a general case of a chi-squared distribution. The Wishart also has the desirable property that random samples of any matrix from this distribution will always be positive definite. This is useful for simulating variance–covariance matrices, which have a positive determinant, and ensuring correlations that lie between −1 and 1.

The variance (or standard deviation as shown in Eq. (5.12)) of the residual uncertainty of the model can be assumed to arise from a uniform distribution (with upper bound b set at an arbitrarily high value), such as

$$\sigma \sim U(0.00001, b) \tag{5.12}$$

Alternatively, the inverse variance of residual uncertainty can be assumed to arise from a gamma distribution,

$$\sigma^{-2} \sim G(a, b), \quad a = b = 0.001 \tag{5.13}$$

The mean of the gamma distribution is given by a/b and the variance by a/b^2. The latter choice of prior has recently been criticized as not being sufficiently uninformative and has been shown, for some examples, to adversely influence the analysis (18).

5.2.3 Biologically Plausible but Low-Information Priors

In PK/PD analyses, there are general boundary conditions on the parameter space which are either biologically illegal (e.g., negative or zero values of clearance) or at the least fairly unlikely (e.g., total blood clearance that is considerably greater than cardiac output). It is not unreasonable, therefore, to consider constructing priors that have their distribution tails (say, the 95% interval) that are somewhere near these "natural" boundary conditions, while at the same time not being overly informative over the remainder of the distribution. These priors could therefore be considered as biologically plausible but weakly informative.

In the section on noninformative priors, a precision of 0.0001 corresponds to a variance of 10,000 (SD = 100), and if it were assumed that the underlying parameter distribution were lognormal (which is common in PK/PD problems), then the 95% interval of the priors would be essentially ~0 and ~+∞. A possible example of a biologically plausible but still low-information prior follows. In this example the values of the parameters are chosen arbtitrarily and any mean values can be used that suit the likely situation. Any choice of mean values will require slight adjustment of the precision matrix; however, this is quite straightforward. For a typical orally administered drug with an assumed fraction absorbed of 1, and mean population parameters for clearance, volume, and absorption rate constant that are the natural log of 1 (L/h), 40 (L), and 1 (h^{-1}), respectively (just over a 24 hour half-life), then the prior could be

$$\boldsymbol{\theta} \sim N(\bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}^{-1}), \quad \bar{\boldsymbol{\mu}} = \begin{bmatrix} \ln(1) \\ \ln(40) \\ \ln(1) \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \tag{5.14}$$

Note that the precision values depend on the likely range of plausible parameter values. For CL and V_d, the precision is set lower than for K_a since the value of the fraction of drug absorbed (F), which scales both parameters, may inflate the apparent values significantly. If F were known to be close to 1, then higher values

of the precision could be used. For this parameterization, the 95% interval for CL, V_d, and K_a are (0.002, 492), (0.08, 19,700), and (0.06, 16), respectively, which for an agent that has approximately these mean values is not overly restrictive. A higher value of precision can be used if there is more than a "vague" expectation about the value of CL from previous studies; for example, CL might be related to renal filtration. In other circumstances, if V_d is expected to be much higher based on previous information, then a larger value of the mean can be used.

Similar to the prior of the mean population parameter values, the prior for the between-subject variance can also be selected to have a more plausible range for PK/PD systems. If we consider the coefficient of variation of between-subject variability for most PK/PD parameters as being approximately <100%, then a choice of ρ for the Wishart distribution that provided a 97.5th percentile value of around this level would be biologically plausible. This is not quite as straightforward as for the precision of the population mean parameter values, since the minimum size of ρ is indexed to the minimum dimension of the variance–covariance matrix of between-subject effects, and ρ affects all variance parameters equally. The value of ρ required to provide a similar level of weak informativeness will vary with the dimension of the matrix. A series of simulations have been performed from the Wishart distribution, where the mean value of the variance of between-subject effects was set at 0.2. The value of ρ required to provide a similar level of weak informativeness will vary with the dimension of the matrix (Table 5.1).

5.2.4 Informative Priors

5.2.4.1 Background of Informative Priors

In many cases the data collected in the current study may be sparse, such as may occur when data are collected as a matter of routine clinical care rather than for model building, which may result in a design that does not support the full model expected based on previous studies. Several options are available to the modeler in these circumstances. One option is to fix the nonestimable parameters to estimates from previous studies. However, the prior estimates may not themselves be sufficiently accurate and therefore inclusion of fixed parameters may lead to biased estimates of the other parameters (19). Another option would be to simplify the model. This, however, may result in a model that is not able to provide useful predictions of future data. A more natural option is to use appropriately informative priors to aid the modeling process.

The setup for informative priors is similar to the concepts provided in Section 5.2.3, regarding biologically plausible but weakly informative priors.

TABLE 5.1   Lowest Value of ρ that Produces a 97.5th Percentile of at Least 100% CV When the Diagonal Elements in Ω_0 Equal 0.2 in the Simulations (Off-Diagonal Components Were 0)

Matrix Dimension	ρ	97.5th Percentile (%CV)
2×2	7	118
3×3	9	106
4×4	11	100
5×5	12	108

One criticism against the use of informative priors is their subjective nature, which may be perceived to introduce bias into the upcoming analysis. The choice of priors and assigning an appropriate level of informativeness is therefore of considerable importance. For population PK/PD studies, there may well be explicit, quantitative data that describes the parameter values in populations that are similar to the population in the current study. In this case it is possible to pool the available information in a meta-analytic technique to provide an appropriate level of prior information. Some care must be taken to assess for heterogeneity between studies and for applicability of studies to the current population under consideration. A brief summary of an approach is shown below. It would be impossible to include an exhaustive treatment of elicitation processes within the confines of this chapter.

5.2.4.2 Inclusion of Studies
The first step is "simply" to find previous published studies that report the PK/PD behavior of interest. Since this process does not have the same goals as a meta-analysis, it is probably not important to include every possible published and unpublished study; however, failing to select a study will introduce subjectiveness and hence potential for bias. Recovering all studies is obviously no mean feat in itself, but general search processes (e.g., MEDLINE) and also the FDA web site (see www.fda.gov) can provide much of the necessary background information. Alternatively, studies may be available on-file and arise from previous clinical or preclinical studies. Information may also be available from other drugs in the same class.

Recording details of the studies, including the models used and associated parameter values reported, is an obvious starting place. Additional details include the chemical analysis method, the pharmacokinetic analysis method, the studied population (specifically subpopulations), number of healthy volunteers or patients, number of pharmacokinetic samples per patient, the dose, the formulation, and the route of administration. If one publication includes several groups of patients (or the same patient received two different formulations/concomitant medications), then each cohort may need to be treated as a repeated measure of the same study or within the same study, which may be indexed according to a study or patient covariate.

In many cases additional work may be required to reparameterize models into the form required for the current analysis. This may involve, for example, a reparameterization between rate constants and clearance and volume terms or between derived parameters, such as volume of distribution by area (V_z) and volume of distribution at steady state (V_{ss}), or even extraction of parameter values from data summary variables (such as peak concentration, C_{max}; time to peak concentration, T_{max}; and area under the concentration curve, AUC). The latter process is sometimes not straightforward and ultimately some data summaries may provide little useful information. See Dansirikul et al. (20) for methods of conversion of data summary variables into model-based parameters.

5.2.4.3 Study Weighting
The heterogeneity of study design among the studies can be treated as a random effect and is ignored unless the studied population differs from the population to be analyzed (e.g., in age and/or renal/hepatic function). For studies where the

populations potentially differ, then either these studies should not be used or additional weighting is required.

For studies that appear to be exchangeable, the weighting is provided by the reciprocal of the estimation variance (which is the precision or Fisher information).

$$w_m = \frac{1}{\text{Var}(\theta_m)} = \frac{1}{SE(\theta_m)^2} \tag{5.15}$$

where w_m is the weight applied to the pharmacokinetic parameter values reported in the mth study (θ_m); and $\text{Var}(\theta_m)$ is the estimation variance of θ_m. However, standard errors (SEs) are often not reported.

In the absence of standard errors, weighting may be applied based on the assumption that the informativeness of the prior study is proportional to the number of subjects (n) in the study. Although this is a rather simplified approach, it can be shown that for designs where the sampling schedule is essentially geometrically spaced and there are more than two times the number of samples as fixed effects parameter values, the approximation will hold (see Duffull et al. (21) for details). Briefly, if we denote the Fisher information matrix for the previous study population as $\mathbf{F}(\theta, \Xi)$, where Ξ represents a given (PK) sampling design for all subjects, then the estimation variance–covariance matrix ($\mathbf{V}$) is given by the inverse of the Fisher information matrix. But in accordance with the Cramér–Rao inequality (readers can refer to Walter and Pronzato (22) for a more in-depth discussion), each diagonal element of the inverse Fisher information matrix is the lower bound of the true but unknown estimation variance,

$$\mathbf{V}_{uu} \geq \left(\mathbf{F}^{-1}(\theta, \Xi)\right)_{uu}, \quad u = 1, \ldots, p \tag{5.16}$$

and the vector of standard errors ($\mathbf{SE}$) of parameter estimates is

$$\mathbf{SE} = \sqrt{\text{diag}(\mathbf{V})} \tag{5.17}$$

where diag denotes the diagonal elements of the matrix. If all subjects in the study receive the same design and have a suitably large number of samples (see above), then

$$\mathbf{F}(\theta, \Xi) = n\mathbf{F}(\theta, \xi) \tag{5.18}$$

where ξ is the design for one subject; therefore, if the information matrix is summarized by the normalised determinant, then

$$\det(\mathbf{F}(\theta, \Xi))^{1/p} \propto n \tag{5.19}$$

That is, a summary measure of the amount of information in the information matrix (given by the normalized determinant) is proportional to the number of individuals in the study, where det denotes the determinant and is a scalar measure of the informativeness of the information matrix. Therefore, it follows that the standard error will be proportional to the inverse of the square root of sample size (Eq. (5.20)), and incorporating Eq. (5.15) provides the approximation

$$SE \propto \sqrt{n^{-1}} = \frac{1}{\sqrt{n}} \qquad (5.20)$$

and the weighting for the mth study by incorporating Eq. (5.20) into Eq. (5.15):

$$w_m = \frac{1}{SE(\theta_m)^2} \approx \frac{1}{\left(\frac{1}{\sqrt{n_m}}\right)^2} = n_m \qquad (5.21)$$

If the number of concentration measurements per subject were smaller than twice the number of fixed effects parameters and the standard errors of the parameter estimates were not provided, then some downweighting would be necessary. It is possible to compute the expected standard errors for any given study (e.g., see Retout et al. (23), but this is beyond the scope of this chapter).

5.2.4.4   Prior Mean for Pharmacokinetic Parameters $\bar{\mu}$ and Ω_0

The priors of the mean population parameters can be computed analogous to how weighted means are computed in meta-analysis (Eq. (5.20)) (24). We show this for a single parameter,

$$\bar{\mu} = \frac{\sum w_m \bar{\theta}_m}{\sum w_m} \qquad (5.22)$$

where $\bar{\mu}$ is an overall mean population parameter, $\bar{\theta}_m$ is a mean population parameter value from the mth study, and w_m is the weighting for the mth study. Similarly, the overall mean between-subject variance (BSV) can be computed for each of the diagonal elements in Ω_0 to produce a matrix of the expected values of the between-subject variance.

5.2.4.5   Computation of the Precision Matrix Σ^{-1}

It is often reasonable to assume that the population parameter values are distributed normally, in which case the pooled estimate of the standard error for a given parameter simplifies to

$$SE = \sqrt{\frac{\frac{\sum_{m=1}^{M} w_m \bar{\theta}_m^2}{\sum w_m} - M\bar{\mu}^2}{M - 1}} \qquad (5.23)$$

where M denotes the number of prior studies.

5.2.4.6   Computation of the Choice of ρ (the Precision of Ω_0)

An empirical method to estimate ρ may be gained by simulation, where candidate matrices of the inverse of Ω are simulated from a Wishart distribution and the empirical distribution of each variance component is compared to the empirical

distribution attained from the prior data. The value of ρ that minimizes the differences of the observed (from prior data) and simulated summary measures (say, quartile ranges) of the distribution of Ω would provide some empirical evidence for the value of ρ.

Overall, there is little literature available that describes formal elicitation processes of parameters from prior studies, while there are several suggested methods for expert elicitation of priors (e.g., see Refs. 25 and 26).

5.2.5   Sensitivity Analysis

Sensitivity analysis is about asking how sensitive your model is to perturbations of assumptions in the underlying variables and structure. Models developed under any platform should be subject to some form of sensitivity analysis. Those constructed under a Bayesian framework may be subject to further sensitivity analysis associated with assumptions that may be made in the specification of the prior information. In general, therefore, a sensitivity analysis will involve some form of perturbation of the priors. There are generally scenarios where this may be important. First, the choice of a noninformative prior could lead to an improper posterior distribution that may be more informative than desired (see Gelman (18) for some discussion on this). Second, the use of informative priors for PK/PD analysis raises the issue of introduction of bias to the posterior parameter estimates for a specified subject group; that is, the prior information may not have been exchangeable with the current data.

The framework of a sensitivity analysis is straightforward in that changes in the posterior distribution are observed under other reasonable prior probability models (27). Sensitivity of posterior to prior distributions can be evaluated by investigation of changes of the posterior distribution to (a) a change in the degree of informativeness of the prior distribution of either itself or other parameters, and (b) a change in the structural form of the prior distribution (e.g., considering the prior as a t rather than a normal distribution). For the former sensitivity analysis, a more weakly informative prior could be chosen by setting the precision of the prior distribution to smaller value(s). For instance, a variety of values for the prior precision for each parameter could be used, for example, where $\sigma^{-2} \rightarrow \frac{1}{2}\sigma^{-2}$, consider $\sigma^{-2} \rightarrow \frac{1}{3}\sigma^{-2}$ or $\sigma^{-2} \rightarrow \frac{1}{4}\sigma^{-2}$ (28).

From a specific PK/PD perspective, where parameters are often assumed to be lognormally distributed, relaxation of this assumption to include a distribution with heavier tails (e.g., a log t-distribution) may be worth considering. This allows for the influence of outliers to be considered explicitly. To accommodate possible outliers, a t-distribution could be used where the degrees of freedom of a student t-distribution can be chosen empirically (29) or estimated during MCMC analysis (30) to provide appropriate weighting to the tails of the distribution.

The differences associated with the sensitivity of the posterior distribution to specification of the prior can then be summarized quantitatively (e.g., % differences of mean and 95% interval) or the distributions can be shown graphically. There are, however, no specific guidelines or criteria available for the assessment of robustness. Any decisions should therefore be made on the basis of the basic question: How does the sensitivity of the posterior to specification of the prior affect the important inferences of the model? This has been addressed in a practical example (31).

5.3   MODEL SELECTION

5.3.1   Approaches to Model Selection

Objective measures of model evaluation are desirable in model building. Ideally, model performance techniques should be indexed to the purpose for which the model is intended to be used and as such should be considered an integral part of all general analysis plans. This brief overview concentrates on methods to select or reject models. By far the majority of methods used for model selection fall within the framework of model discrimination.

Although there are a large number of methods that can be used for model discrimination, we only consider methods that can be implemented easily in WinBUGS (Version 1.4 or earlier) with minimal extra coding.

Model selection criteria are often based on a measure of the fit of the model to the data and a penalty for increased model complexity. The most common method to discriminate between nested models in non-Bayesian population pharmacokinetic analyses is the likelihood ratio test with some predefined level of significance. In fully Bayesian methods there is no "gold standard" model discrimination method, although there are a few methods for dichotomous model discrimination decisions such as the deviance information criterion (DIC) (32). Another relatively common method for model discrimination are the Bayes factors (i.e., the posterior to prior odds ratio). These factors were considered to be the gold standard, although they have also been perceived as overly conservative (27). In essence, they address the problem of how well the prior has predicted the observed data rather than how well the posterior predicts future data and are therefore not defined for models with improper prior distributions. Another, less attractive, feature is that Bayes factors provide the relative probabilities of two models conditional on one of them being true and are therefore most suitable when all candidate models can be specified ahead of time (27, 32). Bayes factors also require some extra coding in WinBUGS and are not considered further here.

Other model selection and/or discrimination tools include the posterior predictive check (PPC) and cross-validation (27, 32, 33). The PPC is useful for examination of the ability of the model to predict accurately certain features of the observed data (e.g., maximum concentration). Although PPC is not strictly a model discrimination technique, as it does not compare the predictive performance between models but rather evaluates the predictive performance of a single model, it does have useful characteristics that are discussed in more detail in Section 5.3.3. Cross-validation is considered accurate but is computer intensive and generally considered not to be suitable for small data sets (32).

More advanced techniques of model evaluation include the reversible jump MCMC (RJMCMC) (34) and "birth–death" algorithms (35). The RJMCMC technique, a form of model averaging, is becoming a popular tool for model selection, which may well become a feature of WinBUGS in the near future. The basic premise involves an MCMC chain that "jumps" to a given model at the current Markov state transition probability. The chain may then jump back, remain, or jump to another model depending on the probability of model preference. It is therefore possible for the chain to explore a number of competing models during the same MCMC run and finally settle on the model of highest posterior probability. The

birth–death algorithm, which is related to the RJMCMC, also allows many models to be compared simultaneously (35).

Further discussion of model selection within the framework of MCMC is divided into discrete model selection, consideration of predictive distributions, and simultaneous modeling of competing models.

5.3.2   Discrete Model Selection

A common form of model selection is to maximize the likelihood that the data arose under the model. For non-Bayesian analysis this is the basis of the likelihood ratio test, where the difference of two $-2LL$ (where LL denotes the log-likelihood) for nested models is assumed to be approximately asymptotically chi-squared distributed. A Bayesian approach—see also the Schwarz criterion (36)—is based on computation of the Bayesian information criterion (BIC), which minimizes the Kullback–Leibler (KL) information (37). The KL information relates to the ratio of the distribution of the data given the model and parameters to the underlying true distribution of the data. The similarity of the KL information expression (Eq. (5.24)) and Bayes's formula (Eq. (5.1)) is easily seen:

$$KL = E\left(\log\left(\frac{\int p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{p(\mathbf{Y}|\boldsymbol{\theta})} \right) \right) \tag{5.24}$$

and the smaller the ratio the closer the distribution of the data under the likelihood approaches the true (but unknown) marginal distribution of the data. This information is, in most real pharmacokinetic problems, notional. The values of the parameters that minimize the KL information will have the highest posterior probability. This result, similar to the non-Bayesian likelihood ratio test, is also based in asymptotic theory (as $n \to \infty$), since at the limit it can be shown that the distribution of the data under the likelihood approaches the true underlying marginal distribution of the data; the information refers to the asymptotic proximity of this ratio. The KL information can also be used to compare two models, where the full model replaces the numerator of Eq. (5.24) and the reduced model replaces the denominator. This method therefore allows the selection of a reduced model that minimizes the difference in a measure of the distribution of the data conditional on the model and parameter estimates and is in reality similar to the non-Bayesian likelihood ratio test.

The Bayesian information criterion (BIC) is defined by

$$BIC = \ln p\left(\mathbf{Y}|\hat{\boldsymbol{\theta}}, M_x\right) - \frac{p}{2}\ln(n) \tag{5.25}$$

where $p(\mathbf{Y} | \hat{\boldsymbol{\theta}}, M_x)$ is the maximum of the likelihood under model M_x and $\hat{\boldsymbol{\theta}}$ a vector of maximum likelihood estimators and p and n carry their previous definitions as the number of parameters (in this case fixed effects parameters only) and number of observations, respectively. The second term (p/2) ln(n) penalizes the BIC function for increasing numbers of model parameters (p). This expression can also be written in the form of $-2LL$ (also termed "deviance" in WinBUGS) as

$$BIC = -2\widehat{LL} + p\ln(n) \tag{5.26}$$

A point estimate of $-2\ \widehat{LL}$ (also termed the deviance (denoted $D(\boldsymbol{\theta})$)) at its maximum is suggested to make the model fit appear better than it should in reality, and in a Bayesian sense averaging over the deviance values (for all values of the posterior distribution of the parameters) would provide a more appropriate choice and so the BIC, now BIC', can be written

$$BIC' = \overline{D(\boldsymbol{\theta})} + p\ln(n) \tag{5.27}$$

The difficulty with using the BIC function (in whatever form) lies in defining the number of model parameters. While this may seem rudimentary, and indeed it is for a linear model, it is not the case for nonlinear hierarchical models. Consider a simple pharmacokinetic example in which a first-order input linear one-compartment model is fitted to the data that arises from 100 patients. The number of model parameters includes: the population parameter values (CL, V_d, K_a), their heterogeneity given by a 3×3 variance–covariance matrix $\boldsymbol{\Omega}$ that has six elements of the lower left triangle, a residual variance component (σ^2), and individual estimates of the parameters ($CL_{1,\ldots,n}{}^T$, $V_{1,\ldots,n}{}^T$, $K_{a1,\ldots,n}{}^T$), which amounts to 310 parameters (in NONMEM only 10 of these parameters enter the approximate population likelihood when using the FO method). Not all of these parameters will contribute equally to the likelihood and indeed some latent parameters may be present. It is therefore not just a matter of summing the parameters to provide a value for p. For nonlinear hierarchical models a deviance information criterion (DIC) has been proposed (32):

$$DIC = \overline{D(\boldsymbol{\theta})} + p_D \tag{5.28}$$

where p_D denotes the number of effective parameters. The number of effective parameters can be calculated as the difference of the posterior mean of the deviance and the deviance at the posterior means of the parameters ($D(\bar{\boldsymbol{\theta}})$),

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) \tag{5.29}$$

and the second term $D(\bar{\boldsymbol{\theta}})$ defines the deviance value for the mean parameter values. This term can be calculated in WinBUGS Version 1.3 by running an additional single iteration after the burn-in period (of at least 4001 iterations, recall that the first 4000 are discarded and cannot be monitored irrespective of chain convergence properties), where the model parameters are fixed at their mean values. Rearrangement of Eqs. (5.28) and (5.29) gives

$$DIC = \overline{2D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) \tag{5.30}$$

Version 1.4 of WinBUGS greatly eases the computational burden by providing DIC as a standard output in the statistical samples toolbox. For users of Version 1.3, the complexity of the model (parameter wise) can more simply be calculated as half of the posterior variance of the deviance as

$$p_D = 0.5\text{Var}(D(\boldsymbol{\theta})) \qquad (5.31)$$

Both of these solutions for the number of expected parameters (Eqs. (5.29) and (5.31)) have been derived from asymptotic theory (27).

Features of using the *DIC* method are essentially its ease of computation and its applicability to nonnested models. Application is simply based on the fact that the model with the lowest *DIC* value is considered best. As a final cautionary note, there are some circumstances when the *DIC* value may provide erroneous results, particularly in circumstances when chain mixing is slow and convergence may not have been achieved (32) (with discussion in Refs. 38 and 39).

5.3.3 Predictive Performance—PPC

In contrast to the hypothesis testing style of model selection/discrimination, the posterior predictive check (PPC) assesses the predictive performance of the model. This approach allows the user to reformulate the model selection decision to be based on how well the model performs. This approach has been described in detail by Gelman et al. (27) and is only briefly discussed here. PPC has been assessed for PK analysis in a non-Bayesian framework by Yano et al. (40). Yano and colleagues also provide a detailed assessment of the choice of test statistics. The more commonly used test statistic is a local feature of the data that has some importance for model predictions; for example, the maximum or minimum concentration might be important for side effects or therapeutic success (see Duffull et al. (6)) and hence constitutes a feature of the data that the model would do well to describe accurately. The PPC can be defined along the lines that *posterior* refers to conditioning of the distribution of the parameters on the observed values of the data, *predictive* refers to the distribution of future unobserved quantities, and *check* refers to how well the predictions reflect the observations (41). This method is used to answer the question: *Does the observed data look plausible under the posterior distribution*? This method is therefore solely a check of internal consistency of the model in question.

It is important to note that PPC does not provide a method for discriminating between models. PPC is included in this section because it does provide evidence for assessment of a given model and therefore has some useful model selection properties. It is possible therefore that a model could be rejected as a possible candidate for describing how the current data arose using a PPC format.

The process is described here. In the following notation we let an observed feature of the data be defined as a statistic denoted by $T(\mathbf{y})$ (a worked through example is provided by Duffull et al. (6)). Simulations of the observed statistic from the posterior distribution of the model predictions are denoted $\mathbf{y}^{\text{rep}}$. We describe the posterior distribution of the model predictions given the observed data as

$$p(\mathbf{y}^{\text{rep}}|\mathbf{y}) = \int p(\mathbf{y}^{\text{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta} \qquad (5.32)$$

which is computed by the integral of the model with respect to the parameters. The probability that the predicted statistic is greater in value than the observed statistic can therefore be written as

$$\Pr(T(\mathbf{y}^{\text{rep}}, \boldsymbol{\theta}) \geq T(\mathbf{y})|\mathbf{y}) \tag{5.33}$$

In this case it is seen that the probability that the predicted statistic $T(\mathbf{y}^{\text{rep}}, \boldsymbol{\theta})$ is greater than the observed statistic $T(\mathbf{y})$ is conditional on the data $\mathbf{y}$. If we assume that the posterior distribution of the parameters in conjunction with the model are a sufficient statistic[1] for the data, then we can write

$$p(\mathbf{y}^{\text{rep}}, \boldsymbol{\theta}|\mathbf{y}) = p(\mathbf{y}^{\text{rep}}|\boldsymbol{\theta}) \tag{5.34}$$

The Bayesian posterior *P-value* can then be computed from

$$P\text{-}value = \int\int I_{T(\mathbf{y}^{\text{rep}},\boldsymbol{\theta})\geq T(\mathbf{y})} p(\boldsymbol{\theta}|\mathbf{y}) p(\mathbf{y}^{\text{rep}}|\boldsymbol{\theta}) \, d\boldsymbol{\theta} \, d\mathbf{y}^{\text{rep}} \tag{5.35}$$

Where $I_{T(.)}$ is an indicator variable that takes the value of 1 if $T(\mathbf{y}^{\text{rep}}, \boldsymbol{\theta})$ is more extreme than or equal to the value of $T(\mathbf{y})$. Although the integral makes the calculation daunting, we can use the power of the MCMC process to provide an approximate numerical solution to the integral. As the number of MCMC iterations (L) approaches infinity, the *P-value* from Eq. (5.34) approaches that from Eq. (5.35).

$$P\text{-}value = \lim_{L\to\infty} L^{-1} \sum_{l=1}^{L} I_{T(\mathbf{y}^{\text{rep}},\boldsymbol{\theta}_l)\geq T(\mathbf{y})} \tag{5.36}$$

An example code for WinBUGS is given in Figure 5.1. The observed test statistic $T(\mathbf{y})$, which is C_{max}, is provided as a comment on line 1. On line 9 the model prediction for the population parameters (mu(1:p)) is assigned to model.pop(j). Here the PPC is performed from the population predictions—rather than the individual predictions. Lines 10 and 11 determine whether the model prediction is greater than the observed test statistic; if so, then a 1 is assigned to the indicator variable Cmax.no(j). Lines 15–17 determine whether any of the predictions for any individual were above the Cmax value (this is computed as the sum of Cmax.ind() > 0) and assigns a 0 to the indicator variable Cmax if any population predictions were above the Cmax, else a 1 is assigned to Cmax. This value is then transformed by subtracting from 1 and assigning to Cmax.ppc. Monitoring the indicator variable Cmax.ppc during the MCMC update will produce summary statistics for PPC.

P-values from the PPC method that are ~0.5 indicate that the model adequately describes the data with approximately 50% of the predictions being more extreme or equal to the observed test statistic. *P*-values close to 0 or 1 indicate some bias in the model predictions and in some circumstances may be used as evidence to reject the candidate model. There is no fixed value of the *P*-value that indicates poor model performance, although values more extreme than 0.1 or 0.9 may confer reasonable evidence against a model.

---

[1] A sufficient statistic in this case is where the model and model parameters are sufficient to describe the distribution of the data and hence it is sufficient to only write the parameters and not show the conditioning of the parameters on the data.

```
 1 # Cmax = 1200 IU/L
 2 model {
 3        for (i in 1:n.ind) {
 4               for (j in off.data[i]:(off.data[i + 1] - 1)) {
 5                      data[j] ~ dnorm(model[j], tau)
 6                      model[j] <- pk.model(1, theta[i, 1:p], time[j], hist[off.hist[i]:(
   off.hist[i + 1] - 1), 1:n.col], pos[j])
 7
 8                      # PPC
 9                      model.pop[j] <- pk.model(1, mu[1:p], time[j], hist[off.hist[i]:(of
   f.hist[i + 1] - 1), 1:n.col], pos[j])
10                      Cmax.int[j]<-max(model.pop[j],1200)
11                      Cmax.no[j]<-equals(Cmax.int[j],model.pop[j])
12               }
13               theta[i, 1:p] ~ dmnorm(mu[1:p], omega.inv[1:p, 1:p])
14        }
15        for (i in 1:n.ind) {
16               Cmax.ind[i]<-sum(Cmax.no[off.data[i]:(off.data[i + 1] - 1)])
17        }
18        # Cmax
19        Cmax.pop<-sum(Cmax.ind[])
20        Cmax<-equals(Cmax.pop,0)
21        Cmax.ppc<- 1-Cmax
22
23        sigma ~ dunif(0, 500)
24        tau <- 1/(sigma*sigma)
25
26        mu[1:q] ~ dmnorm(mu.prior.mean[1:q], mu.prior.precision[1:q, 1:q])
27        omega.inv[1:p, 1:p] ~ dwish(omega.inv.matrix[1:p, 1:p], omega.inv.dof)
28        for (i in 1:p) {
29               for (j in 1:p) {
30                      omega[i, j] <- inverse(omega.inv[1:p, 1:p], i, j)
31               }
32        }
33 }
34
```

FIGURE 5.1   WinBUGS code (Version 1.3) for performing PPC.

5.3.4 Mixture Modeling and Bayesian Model Averaging

Estimation methods that are based on simulation platforms, such as Markov chain Monte Carlo (MCMC), also allow for model discrimination to be based on predictive or posterior distributions. When using MCMC, competing models can be fitted simultaneously as a joint model with an added "mixing" parameter to indicate which model is preferred (42, 43). The posterior distribution of the mixing parameter will provide both the weight of evidence and the posterior probability in favor of one model. The expectation of the prediction from m models and α the mixing parameter can then be evaluated:

$$E[y_{ij}] = \sum_{k=1}^{m} \alpha_k \bar{y}_{ij,M_k} \tag{5.37}$$

where $\bar{y}_{ij,M_k}$ denotes the expectation of the jth observation for the ith subject from model k (M_k) and $\Sigma_{k=1}^{m}\alpha_k = 1$. The likelihood is evaluated:

$$y_{ij} \sim N(E[y_{ij}], \sigma^{-2}) \tag{5.38}$$

In this expression a common residual variance term is specified, although the residual variance can be indexed to the model, in which case the overall residual variance will be the sum of the contribution of the residual variance for each of the m candidate models. It has been found that chain mixing occurs faster when competing models are linked with a common parameter (e.g., the residual error) (42). It is common in the non-Bayesian model framework to address model selection as a

closed decision process—that one and only one model is best. It is, however, also reasonable to consider that no one model under consideration is true and that the data may arise from a multitude of plausible models. The above framework for mixture models does not require that a single model is selected as best (although this assumption will be the case for a worked example). It is also reasonable that all models may be considered as contributing to the likelihood of the data, at the posterior probability of the mixing parameter. It would seem reasonable that models with low probability could be discarded, but there is no particular reason to arrive at a single "best" model.

The mixture model is assessed for a hypothetical example and an example for WinBUGS code is given in Figure 5.2. Evaluation of the individual model predictions for one- and two-compartment models is shown on lines 5 and 7,

```
1 model {
2          for (i in 1:n.ind) {
3                  for (j in off.data[i]:(off.data[i + 1] - 1)) {
4
5                          model1[j] <- pk.model(1, theta.1[i, 1:p1], time[j], hist[off.hist[
   i]:(off.hist[i + 1] - 1), 1:n.col], pos[j])
6
7                          model2[j] <- pk.model(2, theta.2[i, 1:p2], time[j], hist[off.hist[
   i]:(off.hist[i + 1] - 1), 1:n.col], pos[j])
8
9                              model[j] <- model1[j]*(1-mix)+model2[j]*(mix)
10                             var[j] <- model[j]*model[j]*var.cv
11                             tau[j] <-1/var[j]
12                             data[j] ~ dnorm(model[j], tau[j])
13                  }
14                  theta.1[i, 1:p1] ~ dmnorm(theta.mean.1[i, 1:p1], omega.inv.1[1:p1, 1:p1])
15                  theta.mean.1[i, 1] <- mu.1[1]
16                  theta.mean.1[i, 2] <- mu.1[2]
17                  theta.mean.1[i, 3] <- mu.1[3]
18
19                  theta.2[i, 1:p2] ~ dmnorm(theta.mean.2[i, 1:p2], omega.inv.2[1:p2, 1:p2])
20                  theta.mean.2[i, 1] <- mu.2[1]
21                  theta.mean.2[i, 2] <- mu.2[2]
22                  theta.mean.2[i, 3] <- mu.2[3]
23                  theta.mean.2[i, 4] <- mu.2[4]
24                  theta.mean.2[i, 5] <- mu.2[5]
25          }
26
27          sigma ~ dunif(0, 0.5)
28          var.cv<-sigma*sigma
29
30          mix ~ dunif (0, 1)
31
32          mu.1[1:p1] ~ dmnorm(mu.prior.mean.1[1:p1], mu.prior.precision.1[1:p1, 1:p1])
33          for (i in 1:p1){
34                  pop.1[i] <- exp(mu.1[i])
35                  }
36          mu.2[1:p2] ~ dmnorm(mu.prior.mean.2[1:p2], mu.prior.precision.2[1:p2, 1:p2])
37          for (i in 1:p2){
38                  pop.2[i] <- exp(mu.2[i])
39                  }
40          omega.inv.1[1:p1, 1:p1] ~ dwish(omega.inv.matrix.1[1:p1, 1:p1], omega.inv.dof.1)
41          for (i in 1:p1) {
42                  for (j in 1:p1) {
43                          omega.1[i, j] <- inverse(omega.inv.1[1:p1, 1:p1], i, j)
44                          }
45                  }
46          omega.inv.2[1:p2, 1:p2] ~ dwish(omega.inv.matrix.2[1:p2, 1:p2], omega.inv.dof.2)
47          for (i in 1:p2) {
48                  for (j in 1:p2) {
49                          omega.2[i, j] <- inverse(omega.inv.2[1:p2, 1:p2], i, j)
50                          }
51          }
52 }
53
```

FIGURE 5.2   WinBUGS code (Version 1.3) for performing a mixture model.

respectively. The contribution of each model to the expectation of the observation (in accordance with Eq. (5.37)) is given in line 9 with the parameter `mix` representing the fractional contribution of each model at each iteration of the sampler. The likelihood of the data is given in line 12. The second level in the hierarchical model (heterogeneity between subjects) is shown in lines 14–17 for the one-compartment model and lines 19–24 for the two-compartment model. The prior for the residual error (`sigma`—lines 27 and 28) is assumed to be common between both models (this is not an absolute requirement). Lines 32–51 show the one-compartment and two-compartment model priors (note the use of `p1` and `p2` to represent the number of parameters in each model). Monitoring `mix` in the sample toolbox will provide statistics that can be used to support model selection. The expectation of `mix` provides the weight of evidence for the model (clinical importance); the proportion of `mix` >0.5 (or <0.5) provides the statistical evidence for the model.

Due to the limited information that is available describing the use of mixture models, we have exemplified the process in a brief series of simulations. The full analysis is available elsewhere (44). To assess the performance of the mixture model, a hypothetical data set was constructed. Concentrations y_{ij,M_2} at predetermined time points were simulated from a two-compartment (2-c) model with bolus input:

$$E[y_{ij,M_2}] = A_1 \exp(-\lambda_1 t_{ij}) + A_2 \exp(-\lambda_2 t_{ij}) \tag{5.39}$$

The parameters of the model were chosen so that the sum of A_1 and A_2 was 100 ($A_1 = 81.5$, $A_2 = 18.5$) and the ratio of the AUC of the distributional (A_1/λ_1) phase to the total AUC was 0.3, when λ_1 was 1 and λ_2 was 0.1. Parameters for the one-compartment (1-c) model were derived by fitting an exponential equation to data derived from the two-compartment model (without variability or uncertainty):

$$E[y_{ij,M_1}] = A_3 \exp(-\gamma t_{ij}) \tag{5.40}$$

A_3 and γ were estimated at 47.6 and 0.12, respectively. The structural model parameters were derived from the coefficients and exponents assuming a dose of 1000 units. Between-subject variance in the simulation of the data sets was assumed to be lognormal with a value of 0.1 for all parameters. The parameters are presented in Table 5.2. V_1 was restricted to be less than V_2 for all individuals. Sixty data sets

TABLE 5.2   Population Mean and Diagonal Elements of Ω Used in the Simulation of Data Sets

Parameter	Population Mean	Ω
CL (1-c)	2.52	0.1
Vd (1-c)	21	0.1
CL (2-c)	2.81	0.1
V1 (2-c)	10	0.1
Q (2-c)	4.63	0.1
V2 (2-c)	13	0.1

(30 with 1-c profiles and 30 with 2-c profiles) were simulated. Each data set included 20 individuals with 10 observations each (samples drawn after 0.25, 0.5, 1, 2, 4, 6, 12, 24, 36, and 48 h).

An additive residual error was used with a standard deviation of 0.3 units/volume, which was approximately 150–200% of the mean predicted lowest observations. To avoid negative values, a new residual error value was drawn if concentrations were below or equal to zero.

Uncertainty/Precision   The "hypothetical" data sets were run with noninformative priors (flat; precision of 0.0001). The prior parameter means were set to the simulated means. The two competing models were fit simultaneously in WinBUGS Version 1.4 (17) as a mixture model with a mixing population parameter (`mix`—in the following notation `mix` is used in accordance with its use in the WinBUGS code shown in Figure 5.2) drawn from a uniform (0, 1) distribution.

$$E[y_{ij}] = E[y_{ij,M_1}](1 - \texttt{mix}) + E[y_{ij,M_2}]\texttt{mix} \qquad (5.41)$$

The median of the posterior distribution of `mix` close to 0 indicates that the one-compartment model is preferred, while a value close to 1 indicates that the two-compartment model is preferred. Independent residual errors for the two models were used.

The 97.5th percentile of the posterior distribution of the evidence for the true model (1 − `mix` for one-compartment model data and `mix` for two-compartment model data) did not include 0.5 for any of the simulated data sets, indicating that the method selected the right model in all cases. To get reliable estimates on all parameters (including those for the model of low probability), that is, for model averaging, the chains would need to be run for longer than the 10,000 iterations used in this simulation study. In model averaging, the expectation of `mix` and (1 − `mix`) would give the posterior weighted mixture of the one-compartment and two-compartment model, respectively.

5.4   SUMMARY

Bayesian methods offer an attractive framework for the analysis of PK/PD experiments. Previous limitations associated with defining priors, defining models (including as ODEs), defining model selection criteria, and including complex dosing histories (as are common in PK/PD data sets) have essentially become problems of the past. These methods offer flexibility in allowing for previous information to impact on the current analysis, which is an essential part of any sequentially designed drug development program. It is not expected that Bayesian methods, as offered by WinBUGS (or BUGS on other platforms), will supersede current analysis tools—but they will offer a realistic alternative in many circumstances. It is hoped that information provided in the current chapter engages the reader to learn more about the world of MCMC.

REFERENCES

1. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Anal Machine Intelligence* **6**:721–741 (1984).

2. S. Beal and L. B. Sheiner, *NONMEM User's Guide*. University of California, San Francisco, 1992.

3. P. O. Gisleskog, M. O. Karlsson, and S. L. Beal, Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn* **29**:473–505 (2002).

4. L. B. Sheiner, S. Beal, B. Rosenberg, and V. V. Marathe, Forecasting individual pharmacokinetics. *Clin Pharmacol Ther* **26**:294–305 (1979).

5. M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London, 1995.

6. S. B. Duffull, C. M. J. Kirkpatrick, B. Green, and N. H. G. Holford, Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *J Biopharm Stat* **15**:53–73 (2005).

7. D. J. Lunn and L. J. Aarons, Markov chain Monte Carlo techniques for studying interoccasion and intersubject variability: application to pharmacokinetic data. *Appl Stat JRSS Series C* **46**:73–91 (1997).

8. D. G. Fryback, N. K. Stout, and M. A. Rosenberg, An elementary introduction to Bayesian computing using WinBUGS. *Int J Tech Assess* **17**:98–113 (2001).

9. D. J. Lunn, N. Best, A. Thomas, J. Wakefield, and D. Spiegelhalter, Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn* **29**:271–307 (2002).

10. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

11. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. N. Teller, Equation of state calculations by fast computing machines. *J Chem Phys* **21**:1087–1092 (1953).

12. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**:97–109 (1970).

13. W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in Fortran 77*, 2nd ed. Cambridge University Press, Cambridge UK, 1992.

14. A. F. M. Smith and A. E. Gelfand, Bayesian statistics without tears: a sampling–resampling perspective. *Am Statistician* **46**:84–88 (1992).

15. J. Wakefield, An expected loss approach to the design of dosage regimens via sampling-based methods. *The Statistician* **43**:13–29 (1994).

16. L. E. Simonsen, U. Wåhlby, M. Sandstrom, A. Freijs, and M. O. Karlsson, Haematological toxicity following different dosing schedules of 5-fluorouracil and epirubicin in rats. *Anticancer Res* **20**(3A):1519–1525 (2000).

17. D. J. Spiegelhalter, A. Thomas, and N. G. Best, *WinBUGS Version 1.4 User Manual*. Medical Research Council Biostatics Unit, Cambridge, UK, 2003.

18. A. Gelman, *Prior Distributions for Variance Parameters in Hierarchical Models*. `http://www.stat.columbia.edu/~gelman/` (2004).

19. J. R. Wade, A. W. Kelman, C. A. Howie, and B. Whiting, Effect of misspecification of the absorption process on subsequent parameter estimation in population analysis. *J Pharmacokinet Biopharm* **21**:209–222 (1993).

20. C. Dansirikul, M. Choi, and S. B. Duffull, Estimation of pharmacokinetic parameters from non-compartmental variables using Microsoft Excel(R). *Comp Biol Med* **35**:389–403 (2005).

21. S. B. Duffull, T. H. Waterhouse, and J. E. Eccleston, Some considerations on the design of population pharmacokinetic studies. *J Pharmacokinet Pharmacodyn* **32**:441–457 (2005).

22. E. Walter and L. Pronzato, in *Identification of Parametric Models from Experimental Data*, B. W. Dickinson, A. Fettweis, J. L. Masset, J. W. Mdsteine, E. D. Sontag, and M. Thoma (Eds.) Springer, Paris, 1997.

23. S. Retout, S. B. Duffull, and F. Mentré, Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comp Meth Prog Biomed* **65**:141–151 (2001).

24. J. M. Bland and S. M. Kerry, Statistics notes—weighted comparison of means. *Br Med J* **316**(7125):129 (1998).

25. D. Madigan, J. Gavrin, and A. E. Raftery, Eliciting prior information to enhance the predictive performance of Bayesian graphical models. *Comp Stat Theory Method* **24**:2271–2292 (1995).

26. R. A. Jacobs, Methods for combining experts probability assessments. *Neural Comput* **7**:867–888 (1995).

27. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL, 2003.

28. A. Gelman, F. Bois, and J. M. Jiang, Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J Am Stat Assoc* **91**:1400–1412 (1996).

29. J. Wakefield, The Bayesian analysis of population pharmacokinetic models. *J Am Stat Assoc* **91**:62–75 (1996).

30. D. Spiegelhalter, N. Best, W. R. Gilks, and J. Inskip, Hepatitis B: a case study in MCMC methods, in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. Spiegelhalter (Eds.). Chapman & Hall, London, 1996.

31. C. Dansirikul, R. G. Morris, S. E. Tett, and S. B. Duffull, A Bayesian approach for population pharmacokinetic modeling of sirolimus. *Br J Clin Pharmacol* **62**:420–434 (2006).

32. D. J. Spiegelhalter, N. G. Best, B. R. Carlin, and A. van der Linde, Bayesian measures of model complexity and fit. *JRSS Ser B Stat Method* **64**:583–616 (2002).

33. J. B. Kadane, Prime time for Bayes. *Control Clin Trials* **16**:313–318 (1995).

34. P. J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**:711–732 (1995).

35. O. Cappe, C. P. Robert, and T. Ryden, Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *JRSS Ser B Stat Method* **65:**679–700 (2003).

36. G. Schwarz, Estimating dimension of a model. *Ann Stat* **6**:461–464 (1978).

37. S. Kullback and R. A. Leibler, On information and sufficiency. *Ann Math Stat* **22**:76–86 (1951).

38. A. Vehtari and J. Lampinen, *Model Selection Via Predictive Explanatory Power*. Helsinky Laboratory of Computational Engineering, Helsinki University of Technology, 2004, pp. 1–20.

39. G. Celeux, F. Forbes, C. P. Robert, and D. M. Titterington, *Deviance Information Criteria for Missing Data Models*. www.ceremade.dauphine.fr/~xian/cfrt05.pdf (2004), pp. 1–24.

40. Y. Yano, S. L. Beal, and L. B. Sheiner, Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn* **28**:171–192 (2001).

41. D. B. Rubin, Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat* **12**:1151–1172 (1984).

42. B. P. Carlin and S. Chib, Bayesian model choice via Markov chain Monte Carlo methods. *J Roy Stat Soc Br* **57**:473–484 (1995).

43. S. P. Riley, Pharmacokinetic model selection within a population analysis using NONMEM and WinBUGS, in *AAPS Workshop on Bayesian Primer*. AAPS, Salt Lake City, UT, 2003.

44. L. E. Friberg, C. Dansirikul, and S. B. Duffull, Simultaneous fit of competing models as a model discrimination tool in Bayesian analysis, in *PAGE Meeting Uppsala*. `http://www.page-meeting.org/page/page2004/posters.pdf` (2004).

Estimating the Dynamics of Drug Regimen Compliance

ENE I. ETTE and ALAA AHMAD

6.1 INTRODUCTION

A major problem in pharmacotherapy is noncompliance with prescribed medication regimens (1). Noncompliance has been used to indicate a range of components of nonadherence with assigned treatment. Urquhart (2, 3) distinguishes noncompliance into three phases. (1) *Acclimatization period*: This is the period a patient considers the acceptance of the concept of treatment and his/her agreement to execute the prescribed regimen. Although the patient has the formal right to reject the agreement, some do so covertly by not starting to take the prescribed medicine. Some patients may even open a pill container a few times the first week and may never open it again, making the decision process take a while. This, however, should not be confused with the delay to start taking the medication. A costly error can occur if nonacceptance is not recognized, resulting in the interpretation of nonresponse as drug-refractory disease. (2) *Compliance with the decision*: During this phase patients who accept to start drug therapy implement the prescribed regimen in a more or less punctual manner. The accepted treatment is experienced as a continual process, and with it opportunities for errors in timing and dosing as well. (3) *Discontinuation*: This is an abrupt end to the previous phase. It occurs when a patient discontinues treatment of his/her own accord or on medical advice. This is a common occurrence that should not be confounded with long drug holidays.

Drug holiday refers to interruption in dosing for a period of three or more days (4). An arbitrary duration is used in the definition of drug holidays. Consequently, it can have different effects depending on the disposition pharmacokinetics of the drug and regimen. For a drug with linear/dose proportional pharmacokinetics, the impact of missed doses on the concentration following a missed dose can be expressed as $n(\tau/t_{1/2})$, where τ is the prescribed dosing interval, $t_{1/2}$ is the elimination half-life, n is the number of doses missed, and $\tau/t_{1/2}$ is the noncompliance impact

factor (5). The impact of an omitted dose on concentration is less pronounced if the noncompliance impact factor is small. In addition, pharmacodynamics should be considered in the definition to reflect meaningful clinical effect. Levy (5) introduced the notion of therapeutic coverage (maintenance of effective plasma concentration)—the ratio of the time to reach the minimal effective concentration divided by τ—to deal with this.

Noncompliance is used in the remainder of the chapter interchangeably with nonadherence (3) to cover the different phases of noncompliance described above. Thus, medication noncompliance can simply be defined as not adhering to physician instructions concerning prescription medications. Some examples of noncompliance include not having a prescription filled or refilled, taking too much or too little of the medication, erratic dosing due to forgetfulness, and discontinuing the medication too soon.

The importance and impact of noncompliance has been cited in a government report. A US senate subcommittee study on medication noncompliance documented that not taking medications as directed results in over 300,000 deaths in the United States annually, and 125,000 deaths in recovering cardiac patients alone (6).

6.2   MEASUREMENT OF COMPLIANCE

Several methods have been used to measure patient compliance to drug therapy. Some of these methods lack the sensitivity to detect individuals who truly do not take the drugs prescribed (7–10). They include direct questioning and the use of interview instruments, patient diaries, and pill counts. The pill count approach tends to overestimate adherence (11–14), often due to "medication dumping" (15). Patient self-report, which has been used extensively to assess compliance, also tends to overestimate adherence (16, 17). Although drug levels are an objective measure of drug exposure, they provide only a snapshot of behavior and are affected by factors other than adherence (18). A more direct way to confirm drug ingestion is the incorporation of a chemical marker into the dosage form and qualitatively detecting its presence in a biological fluid (usually plasma or urine) (19). The incorporation of a chemical marker may not reveal the extent of drug ingestion such as underdosing or overdosing. The Medication Event Monitoring System (MEMS) (20–22) is a relatively recent method that provides an objective measure of pill bottle opening and use of an inhaler spray or other applicator, but not necessarily pill taking or inhalation from the inhaler (23), and may underestimate compliance. No method for assessing compliance is completely accurate.

These methods of measuring compliance continue to be used in clinical trials, although none of them fulfill the criterion of providing an accurate measurement of drug taken. Since they are not, in and of themselves, acceptable measures of compliance, they can only serve to confound the analysis and interpretation of the compliance exposure–response relationships. However, ancillary information from patient diaries, for example, may be helpful in interpreting data collected by other means—such as MEMS and others described above (24). A combination of MEMS and patient diaries may give an unbiased estimate of compliance.

6.3   COMPLIANCE INDICES

Various aspects of drug-taking behavior are quantified with compliance indices. The fraction of doses taken in the monitoring period, which is analogous to a pill count, and the fraction of days during which the patient adhered to the prescribed dosing frequency are the most commonly reported compliance indices (10, 25–30). Dose timing has been measured as the fraction of doses taken at the prescribed dosing intervals (10, 25, 28, 30). More often than not an appropriate grace period (20–25%) is allowed for the latter. Another index of compliance measurement is "therapeutic coverage," which was discussed in the Section 6.1. In addition, Ahmad et al. (31) introduced "time at risk"—the duration of time when subjects have subtherapeutic concentrations and may be at risk of developing breakthrough symptoms—as a compliance index for monitoring compliance to antiepileptic therapy.

6.4   PROBABILITY BASIS OF COMPLIANCE

6.4.1   Markov Chain

A series of probable transitions between states can be described with the Markov chain. A Markovian stochastic process is memoryless, and this is illustrated subsequently. We generate a sequence of random variables, $(y_0, y_1, y_2, \ldots)$, so that each time $t \geq 0$, the next state y_{t+1} would be sampled from a distribution $P(y_{t+1}|y_t)$, which would depend only on the current state of the chain, y_t. Thus, given y_t the next state y_{t+1} would not depend additionally on the history of the chain $(y_0, y_1, y_2, \ldots, y_{t-1})$. The name *Markov chain* is used to describe this sequence, and the *transition kernel* of the chain is $P(.|.)$. $P(.|.)$ does not depend on t if we assume that the chain is time homogeneous. A detail description of the Markov model is provided in Chapter 26.

   In considering how the initial state of y_0 impacts y_t, the distribution of y_t given y_0, denoted here as $P^{(t)}(y_t|y_0)$, needs to be examined. y_t depends directly on y_0 because the intervening variables $(y_0, y_1, \ldots, y_{t-1})$ are not provided. $P^{(t)}(.|y_0)$ will eventually converge to a unique invariant (or stationary) distribution that is independent of y_0 or t as the chain gradually "forgets" its initial state, subject to regularity conditions.

6.4.2   Model for Medication Compliance

Compliance comprises taking the drug at the prescribed dose—*dosage compliance*—and taking the drug at the scheduled times—*dosing time compliance*. Although it has been argued that the indices of dosage compliance are usually less variable than that of dosing time compliance (32), this aspect of compliance should not be ignored. Separate modeling may be required for each of these medication errors (see Wang (33)). Figure 6.1 illustrates the two related processes in compliance.

   Girard et al. (34) proposed a hierarchical Markov model for patient compliance with oral medications that was conditioned on a set of individual-specific nominal daily dose times. The individual random effects for the model were assumed to be multivariate normally distributed. Assuming first–order Markov hypothesis (see
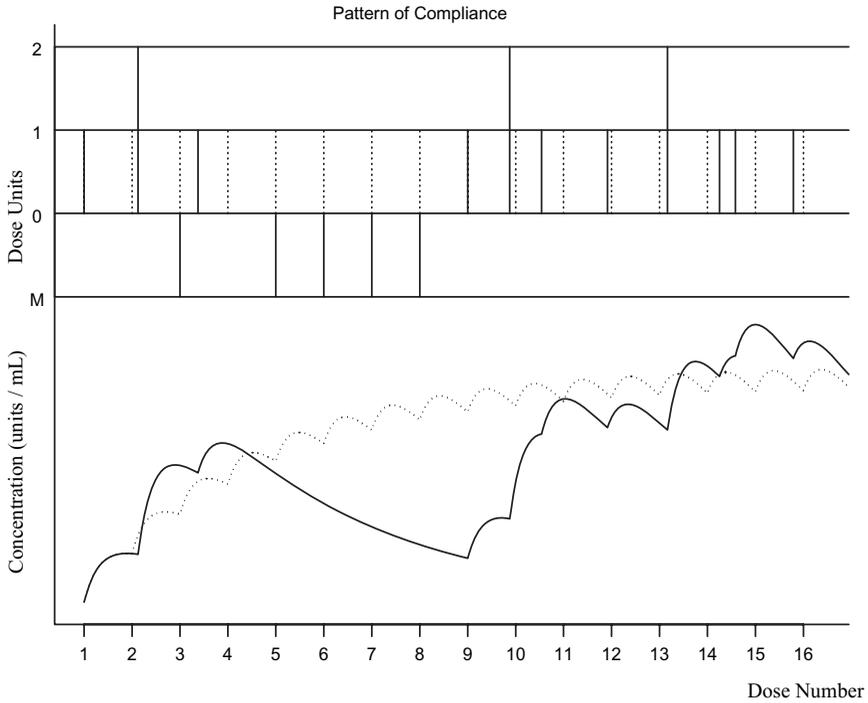
FIGURE 6.1   A simulated nonadherence concentration profile for a subject over 16 nominal doses. The dosing pattern for the patient is described in the upper half of the figure. Each solid bar represents an actual dosing time, and each dotted bar a scheduled dosing time. The height of the solid bar can be 1 – one dose taken, 2 – two doses taken, and 1 (i.e., M in the figure)—a dose missed or not taken.

Chapter 26), the probability of a subject taking a dose or doses (one or more) at any given dosing time is a function of whatever doses were taken at the immediate past dosing time preceding the one in question. This, of course, is independent of all other previous dosing events—a Markov property.

   If a subject did not take his/her medication given that it was not taken the time before, or did take the medication given that it was taken the previous dosing time, a two-state Markov chain model can be fully defined by using two conditional probabilities. By defining $Y = (y_1, y_2, \ldots, y_n)$ as a random vector indicating whether a patient/subject has not taken his/her medication ($y_i = NT$) or has taken it ($y_i = T$) at ith time, then

$$p(y_i = NT \,|\, y_{i-1} = NT) = P_{00} \tag{6.1}$$

$$p(y_i = T \,|\, y_{i-1} = T) = P_{11} \tag{6.2}$$

describe the probabilities of these events (34).

   From Eq. (6.1) and (6.2),

$$p(y_i = T \mid y_{i-1} = NT) = P_{01} = 1 - P_{00} \tag{6.3}$$

$$p(y_i = NT \mid y_{i-1} = T) = P_{10} = 1 - P_{11} \tag{6.4}$$

The above probabilities take on values between 0 and 1 and vary from individual to individual instead of having a 50/50 probability that would be expected for an unbiased coin. The Markov chains described above are sometimes called "two-coin" models, corresponding to a subject having two virtual coins that could be tossed alternately. The alternate tossing would be dependent on whether a dose was taken or not taken. Using a mixed effects logistic regression, the interindividual probability for a subject who may continue in a study, for instance, without taking the medication can be modeled as follows:

$$P_{00} = \exp(\alpha_1 + \beta_1 Z + \eta_1) / [1 + \exp(\alpha_1 + \beta_1 Z + \eta_1)] \tag{6.5}$$

where α_1 is the intercept, β_1 is a vector of covariate parameters, Z is a matrix of individual covariates, and η_1 is a random effect parameter, with mean 0 and variance Ω. A similar model can be written for the probability P_{11} that a subject may stop taking the medication. A detailed description of Markov mixed effect regression modeling for compliance data can be found in Girard et al. (34).

6.4.3   Other Methods for Handling Medication Compliance Data

Other approaches to modeling medication compliance data have been reported in the literature and are described briefly below.

6.4.3.1   Random Sampling Probabilistic Model Approach
Hughes and Waley (35) described a probabilistic model that was used to characterize dose-taking behavior of subjects (patients) in a lipid-lowering agent study. They used a random sampling of adherence patterns to drive the model that described the onset and offset of drug effects. Patients could either comply (with a probability P) or not comply (with a probability $1-P$) when faced with their first dose. The probability of taking the next dose decreased as a function of time if the dose was missed as follows:

$$P_{(\text{taking\_next\_dose})} = P \exp^{-kT_{\text{off}}} \tag{6.6}$$

where k is the rate constant determining the time-dependent decrease in the probability of taking a dose T_{off} days after a previous dose was taken. T_{off}, therefore, is time since the last dose was taken. T_{off} is 0 and $P_{(\text{taking\_next\_dose})}$ is P once again when dosing is resumed.

Drug holidays and premature discontinuation can be modeled with the same expression because $P_{(\text{taking\_next\_dose})}$ is time dependent. With the model in Eq. (6.6), a prolonged drug holiday eventually leads to discontinuation because the model assumes that dose-taking compliance and persistence are related.

6.4.3.2   Likelihood Approach
A likelihood approach for selecting from a set of possible dosing histories for each individual in a data set when such an individual has more than one dosing history

available was proposed by Jonsson et al. (36). The population model developed was based on individuals whose dosing histories were very similar.

6.4.3.3   Bayesian Objective Function Approach

A Bayesian objective function has been used to describe observed data collected after a given dose based on three different patient behaviors regarding the dose (37). The patient medication-taking behaviors were categorized as (a) dose was not taken, (b) dose was taken at the correct time, and (c) dose was taken but at an undetermined time. Using simulations, the authors showed that the approach provided a set of pharmacokinetic (PK) observations that performed well for subsequent exploratory analyses or estimating individual parameters. The approach, however, is limited by the fact that the study drug had to be a drug with a short elimination half-life.

6.4.3.4   Hierarchical Bayesian Approach

Mu and Ludden (38) described a hierarchical Bayesian model-based approach to incorporate an estimate of compliance into a population PK analysis. With the approach, both compliance and population PK parameters were estimated simultaneously. Rather than emphasize the estimation of compliance, the emphasis of the approach is on improving efficiency in the estimation of population PK parameters.

6.4.3.5   Missing Dosing History Approach

Soy and colleagues (39) used a simulation study to investigate some methods in which past dosing history (including nonmissing dose information) was "ignored." The data were simulated such that the pretest dose for all individuals had a "dose timing error" as well as a possible "dose amount error." They compared different assumptions about the amounts in the compartments at time zero. The assumptions were categorized into (a) prescribed dose method (PDM), (b) missing dose method (MDM), (c) missing dose mixture method (MDMM), and (d) extrapolation–subtraction method (ESM).

The PDM was consistent with the intention-to-treat principle (40). An intention-to-analysis is one in which the analyst essentially discards all information about compliance to assigned treatment.

Dosing history was not taken into account with the MDM approach. A new parameter (A_0), corresponding to the (unknown) amount of drug in the central compartment at time 0, was estimated in addition to population PK parameters. It was modeled as an individual's value of a generic PK parameter:

$$P = \mu_P \, \exp(\eta_P) \tag{6.7}$$

where the typical value of the parameter in the population (P) is μ_P, and η_P is a normally distributed random effect (interindividual variability) with mean 0 and variance Ω_P. The implicit assumption made with the MDM is that the amount of drug at time 0 in the depot compartment is zero.

Individuals are assumed to arise randomly from one of two subpopulations with MDMM. Subjects with PK data compatible with the nominal dose history comprise

subpopulation 1, while subjects with PK data inconsistent with nominal dosage comprise subpopulation 2. The latter are subjects for whom the MDM is suitable. Estimated simultaneously with the population PK parameters is the probability of belonging to one subpopulation or another.

The log of the slope (K) of the terminal-elimination phase of an observed time–concentration curve for an individual (obtained from the last three to four PK observations) was used to build a "subtraction curve" between time $t = 0$ and $t = \tau$ equal to $C_0 \exp(-Kt)$ with ESM . The value of this curve was subtracted from the "true" observation, at the time of each actual observation, and these differences were analyzed as though they were the observed PK responses to a single dose. The authors included this method to imitate what is often done in practice when "baseline" concentrations are clearly nonzero and prior dosing history is deemed to be unreliable or unavailable.

Thus, all dosage history is completely ignored with MDM when this history is sufficiently suspect. On the other hand, with MDMM probabilities were assigned to the events where in one case an individual's reported history could be used to describe the subject's data, or in another case in describing a subject's data this history could be ignored. These probabilities were then used to appropriately downweight past dosing history (39). The results of the simulation study indicated that there was little basis for preferring MDMM over MDM (39). However, when all dosing histories are at least somewhat wrong, the two-subpopulation mixture model (MDMM) cannot add great value, and the authors concluded that the possible benefit of MDMM was more theoretic than real.

In practice, some dosing histories may be accurate and it would seem counterproductive to deprive an analysis of all valid dosing histories just because some are invalid. Although the authors assumed a missing data scenario where no doses were assumed to be unreliable other than the one immediately preceding the test dose, neither MDM, MDMM, nor ESM used any prior dosing history whatsoever and they performed similarly in parameter estimation efficiency. Thus, each of these approaches offered similar performance advantage over PDM. Therefore, the choice of the method to use and the ease with which a particular approach can be implemented is left to the pharmacometrician. It is worth noting, however, that when previous dosing history is unavailable, MDM, a method that does not rely on such a history, would be a preferred choice for estimating population PK parameters rather than PDM, which assumes perfect compliance.

6.4.3.6   *Maximum Penalized Marginal Likelihood (MPML) Approach*

Kenna and Sheiner (41) used a simulation study to show that the MPML method—which uses an all compliance data–dosage history questionnaire, C_Q, available from all subjects and combines that with dosing history obtained with MEMS, C, from a random fraction of subjects, effectively calibrating C_Q to C—is superior to other methods that use only one compliance measure, or both, or neither; where neither was intention-to-treat. The authors showed that the MPML approach yielded efficient dose–response estimates over a wide range of clinical trial designs, effect sizes, and varying quality and quantity of compliance information. The method was shown to maintain good performance even when its key assumptions were violated and compliance data were sparse.

6.5   NONCOMPLIANCE AND STEADY-STATE PHARMACOKINETICS

When patients comply well with prescribed drug therapy, the combination of a steady-state assumption and knowledge of the time of the last dose has proved sufficient for the extraction of knowledge from plasma concentration measures taken a few times from each subject in a study population. As discussed in Section 6.1, patients often comply poorly with medication regimens, even when they take the prescribed number of doses. When subjects are noncompliant, the assumption of steady-state pharmacokinetics or steady state prior to the last dose taken does not hold and results in an erroneous evaluation of the history of drug exposure and consequently biased estimates of PK parameters (42–44). Vrijens et al. (45) emphasize the importance of combining MEMS data with PK data to estimate PK parameters without making steady assumptions. Thus, it is of the utmost importance to integrate appropriately collected compliance data, such as that obtained with MEMS or MEMS with patient diaries, in analyzing PK data. Biased PK parameter estimates that are used to drive a pharmacokinetic (PK)/pharmacodynamic (PD) model would affect the outcome of such an analysis. Implications of integrating adherence, adherence modeling in therapeutics, and drug development have been reviewed in the literature (32, 46).

6.6   APPLICATION

In this section we use a simulation study systematically to characterize the effect of noncompliance on steady-state pharmacokinetics. Specifically, the effect of missed and replacement doses on the steady-state pharmacokinetics of valproic acid (VPA) following the ER and DR preparations of the drug were investigated (31).

Divalproex sodium extended-release (Depakote® ER) is a once daily (QD) formulation for VPA that was developed to improve patient compliance and reduce side effects compared to the standard twice-daily (BID) delayed release (DR) formulation (Depakote® tablets). However, there are concerns of potential subtherapeutic concentrations following delayed or missed doses or toxic concentrations with replacement doses for the ER and DR formulations.

6.6.1   Simulation

A one-compartment model with first-order elimination was used to simulate unbound VPA concentrations. The two formulations differ only in the input function: the ER formulation was accounted for through a zero-order input over 22 hours with 89% bioavailability. The DR formulation absorption was characterized by a 2 h lag time ($t_{lag} = 2$ h) followed by first-order absorption rate ($k_a = 0.1\,h^{-1}$). The bioavailability (F) of the DR preparation was assumed to be complete ($F = 1$).

Equation (6.8) was used to simulate unbound VPA concentrations (C_u) following administration of the DR preparation and Eq. (6.9) was used to simulate C_u following the ER formulation:

$$C_u = \left(k_a D/V\left(k_a - CL_u/V_u\right)\left(e^{-CL_u/V_u t} - e^{-k_a t}\right)\right) \tag{6.8}$$

$$C_u = FD/CL_u \cdot \mathrm{T}\left(e^{CL_u/V_u T} - 1\right)e^{-CL_u/V_u t} \tag{6.9}$$

where D is the dose, V_u is the volume of distribution of unbound drug, CL_u is the systemic clearance of unbound drug, F is bioavailability, k_a is the first-order absorption rate constant for DR, and T is the duration of the zero-order input for ER.

The following equation was used to calculate the total VPA concentrations (C_t):

$$C_t = C_u + (N_1 K_1 C_u P)/(1 + K_1 C_1) + (N_2 K_2 C_u P)/(1 + K_2 C_u) \qquad (6.10)$$

where P is albumin concentration, N_1 and K_1 and N_2 and K_2 are the number of binding sites and equilibrium association constants for a low-affinity–high-capacity binding site and high-affinity–low-capacity binding site, respectively.

Each simulation included 100 hypothetical subjects. The model parameters used were derived from an adult population and there were no covariate distribution models for the virtual trial population. Subjects were assumed to be healthy and on valproate monotherapy (31). The simulations assumed that the extended release (ER) formulation was administered once daily and the delayed release (DR) preparation was administered twice daily. Unbound and total valproic acid concentrations were simulated from the time of dose administration to 280 h; and the simulations were based on the administration of 1000 mg ER once daily, 500 mg DR twice daily, 2500 mg ER once daily, and 1000 mg DR twice daily. For once-daily regimens, simulation scenarios included doses taken 6, 12, 18, and 24 h late from schedule and then two doses taken 24 h late (replacement dose for the missed dose). For the twice-daily regimens, doses were simulated 3, 6, 9, and 12 h later than the scheduled times and then two doses were simulated 12 h later than scheduled to mimic replacement dosing for a missed dose. More extreme cases where two doses are delayed at various times or missed were also simulated.

The following parameters (geometric means ± SD) for CL_u and V_u were assumed: 5.04 ± 1.00 L/h and 95.1 ± 19.0 L, respectively. These parameters were assumed to be lognormally distributed. Protein binding parameters, $N_1 = 1.54 ± 0.108$, $K_1 = 11.9 ± 1.99\,\text{mM}^{-1}$, $N_2 = 0.194 ± 0.0783$, $K_2 = 164 ± 141\,\text{mM}^{-1}$, and $P = 0.528 ± 0.0528\,\text{mM}$, were assumed to be normally distributed. Limits of ±2 standard deviations were placed on all parameters for the simulations.

VPA concentration versus time profiles were generated for each scenario. Drug concentrations were compared to the therapeutic range of valproic acid. Based on total VPA, a therapeutic range of 50–150 mg/L was assumed. The lower limit for the therapeutic range for unbound VPA was 5 mg/L. (At total concentrations of 50 mg/L, almost 90% of the binding sites are occupied; therefore, the free fraction is 10%.) There is no accepted upper limit for the therapeutic range of unbound VPA.

6.6.2   Data Analysis

In order to assess the effect of missed or delayed doses, the simulation outcomes were summarized by:

1. Number of subjects with subtherapeutic concentrations after delayed or missed doses quantified as the percentage of subjects having total drug concentrations lower than 50 mg/L or unbound VPA concentrations less than 5 mg/L. Subtherapeutic subjects at baseline steady state were excluded from poor adherence scenarios.

2. The time range that subjects spent in the subtherapeutic range or "time at risk," in hours, which is essentially the duration of time where subjects might be at risk of breakthrough symptoms.

3. Number of subjects with drug concentrations above the upper limit of the therapeutic range quantified as the percentage of subjects with total VPA concentrations exceeding 150 mg/L. This percentage reflects the probability of potential toxicity.

6.6.3 Results

Figures 6.2 and 6.3 are examples of disrupted PK profiles as a result of noncompliance to the ER preparation once-daily regimen. The simulated total valproic acid concentrations following administration of 2500 mg daily of the ER preparation are shown in Figure 6.2. ER dose on day 7 was administered 6 h late from schedule (30 h after the last dose on day 6). The effect of the missed dose followed by a doubling of the dose is shown in Figure 6.3.

The percentage of subjects on ER 1000 mg who had subtherapeutic concentrations due to poor compliance varied from 43% to 100% with respect to C_u (<5 mg/L) and from 28% to 100% with respect to C_{tot} (<50 mg/L) (see Table 6.1) The mean "times at risk" varied from 6 to 60 h with respect to C_u and from 8 to 53 h with respect to C_{tot} (see Table 6.1).

None of the subjects on ER 2500 mg QD had subtherapeutic concentrations even if one dose was delayed 6 h from schedule. Almost 50% of the population had subtherapeutic concentrations if one dose (ER 2500 mg) was missed from schedule while all subjects would be subtherapeutic if two doses were missed. The mean "time at risk" varied from 0 to 28 h (Table 6.1). Regarding potential toxicity ($C_{tot} > 150$ mg/L), 52% of the population would experience toxic concentrations if two doses were taken 66 h after last dose while on ER 2500 mg QD.
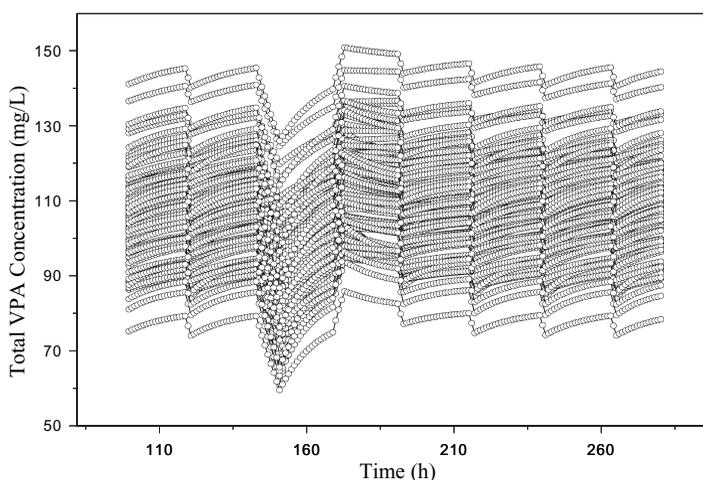

FIGURE 6.2 VPA levels following administration of 2500 mg daily. Dose on day 7 was administered 6 h late.
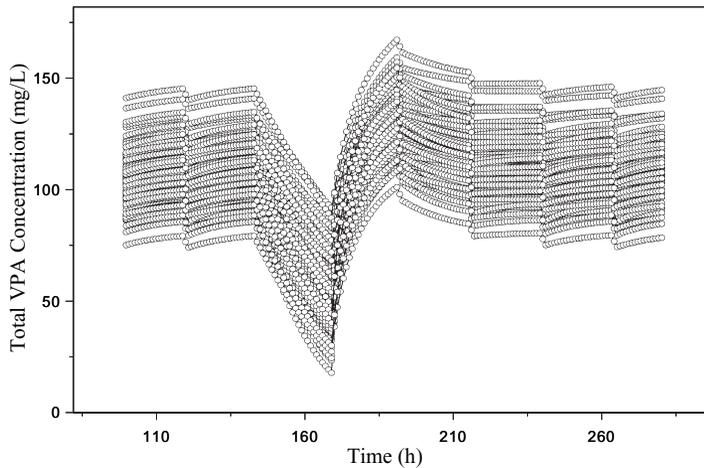
FIGURE 6.3   VPA levels following administration of 2500 mg daily. Dose on day 7 was missed. Dose was doubled on day 8.

For the DR 500 mg BID regimen, the percentage of subjects who had subtherapeutic concentrations due to nonadherence varied from 3% to 88% with respect to C_u and from 3% to 77% with respect to C_{tot}. The mean time at risk varied from 1 to 14 h (Table 6.2). None of the subjects experienced subtherapeutic concentrations if one dose was delayed or missed from a DR 1000 mg BID regimen. However, if two doses are delayed from schedule, 1–24% of the population would have subtherapeutic concentrations. The mean "time at risk" varied from 2 to 6 h. Dosing recommendations following missed doses of ER and DR VPA formulations are shown in Tables 6.3 and 6.4, respectively.

6.6.4   Summary of the Simulated Study Findings

Higher doses of the ER preparation (2500 mg QD) could be used to provide adequate seizure control with dose delays up to 12 h. For unstable seizure patients, it is recommended that patients maintain a twice-daily regimen since twice-daily regimens are less susceptible to fluctuations in steady-state concentrations in the case of noncompliance. For a shorter dosing interval, twice-daily regimens demonstrate better maintenance of drug concentrations in the case of delayed or missed doses.

6.7   SUMMARY

Noncompliance with prescribed medication regimens is a major problem in pharmacotherapy and results in 300,000 deaths in the United States annually. There are three distinguishing phases to noncompliance: (a) acclimatization period, (b) compliance with the decision, and (c) discontinuation. Several methods are used to measure patient compliance to drug therapy. They range from direct questioning and the use of interview instruments, to patient diaries, pill counts, MEMS, drug

TABLE 6.1 Summary of Simulation Scenarios for Once-Daily ER Regimen

| | C_u | | | C_{tot} | | | |
	% Subtherapeutic (<5 mg/L)	Time at Risk (hours) Range	Time at Risk (hours) Mean	% Subtherapeutic (<50 mg/L)	Time at Risk (hours) Range	Time at Risk (hours) Mean	% Subjects with C_{tot} > 150 mg/L
Scenario							
ER 1000mg QD							
One dose taken x hours after last dose							
30	43	1–20	6	28	1–24	8	0
36	78	1–25	13	65	1–26	12	0
42	90	1–27	18	82	2–29	16	0
48, missed dose	96	1–53	28	81	1–52	26	0
48, two doses	96	1–29	20	83	1–32	19	0
Two doses taken x hours after last dose							
54	100	7–37	28	100	1–40	25	0
60	100	14–44	34	100	6–47	31	0
66	100	21–48	40	100	13–50	37	0
72, two doses	100	28–54	46	100	9–59	43	0
72, one dose	100	39–83	60	100	19–86	53	0
ER 2500mg QD							
One dose taken x hours after last dose							
30	0	0	0	0	0	0	1
36	3	1–3		5	2–4	3	2
42	20	1–11	5	16	1–12	6	5
48, missed dose	46	2–17	8	47	1–24	8	0
48, two doses	51	1–17	7	46	1–17	7	8
Two doses taken x hours after last dose							
54	73	1–23	11	68	1–23	11	14
60	84	1–29	12	80	3–30	15	32
66	91	1–35	21	88	3–37	20	52
72, two doses	91	1–41	26	92	1–46	25	3
72, one dose	100	1–42	28	95	3–43	27	0

TABLE 6.2 Summary of Simulation Scenarios for Twice-Daily DR Regimen

Scenario	C_u			C_{tot}		
	% Subtherapeutic (<5 mg/L)	Time at Risk (hours) Range	Mean	% Subtherapeutic (<50 mg/L)	Time at Risk (hours) Range	Mean
DR 500 mg BID						
One dose taken *x* hours after last dose						
15	3	N/A	1	3	1–4	2.5
18	11	1–7	4	10	1–12	5
21	25	1–10	5	21	1–13	5
24, missed dose	39	1–33	12	34	1–37	12
24, two doses	39	1–13	6	29	1–16	7
Two doses taken *x* hours after last dose						
27	64	1–17	7	43	1–21	8
30	76	1–21	10	61	1–25	9
33	90	1–23	12	75	1–26	11
36	88	2–29	14	77	1–31	14
DR 1000 mg BID						
One dose taken *x* hours after last dose						
15	0	N/A	0	0	N/A	0
18	0	N/A	0	0	N/A	0
21	0	N/A	0	0	N/A	0
24, missed dose	0	N/A	0	0	N/A	0
24, two doses	0	N/A	0	0	N/A	0
Two doses taken *x* hours after last dose						
27	1	2	2	4	2–3	2.2
30	5	1–6	3	19	1–8	4
33	19	1–9	3	15	1–11	5
36	24	1–11	6	22	1–13	6

TABLE 6.3   Dosing Recommendations for ER Based on Free VPA

ER 1000 mg QD	
One dose taken *x* hours after last dose	
30	Take dose and resume dosing
36	Take dose and resume dosing
42	Take dose and resume dosing
48, missed dose	Take make-up dose
48, two doses	Take doses and resume dosing
Two doses taken *x* hours after last dose	
54	Take doses and resume dosing
60	Take doses and resume dosing
66	Take doses and resume dosing
72, two doses	Take doses and resume dosing
72, one dose	Take two doses and resume
ER 2500 mg QD	
One dose taken *x* hours after last dose	
30	Take dose and resume dosing
36	Take dose and resume dosing
42	Take dose and resume dosing
48, missed dose	Take dose and resume dosing
48, two doses	Do not double the dose
Two doses taken *x* hours after last dose	
54	Risk of toxicity, take 1.5 dose
60	Risk of toxicity, take 1.5 dose
66	Risk of toxicity, take 1.5 dose
72, two doses	Take doses and resume
72, one dose	Take two doses and resume

concentrations, and chemical markers. The pill count and patient self-report tend to overestimate adherence. MEMS may underestimate compliance, but combining it with patient diary may provide unbiased estimate of compliance. The incorporation of a chemical marker may not reveal the extent of drug ingestion such as underdosing or overdosing.

Several indices are used to quantify compliance. These range from fraction of doses taken in the monitoring period (i.e., a pill count) to the fraction of days during which the patient adhered to the prescribed dosing. Dose timing has been measured as the fraction of doses taken at the prescribed dosing intervals. Other indices of compliance measurement are "therapeutic coverage" and "time at risk."

Probabilistic models have been developed for characterizing compliance. The most commonly cited probabilistic approach is the hierarchical Markov model. Other more recently developed approaches range from a random sampling probabilistic model approach, to likelihood approaches, Bayesian approaches, and a missing dosing history approach. It is up to the pharmacometrician to choose the method that would best characterize his/her nonadherence data. The application example reinforces the importance of compliance to prescribed drug therapy, and how steady-state pharmacokinetics can be disrupted in the presence of noncompliance.

REFERENCES

TABLE 6.4   Dosing Recommendations for DR Based on Free VPA

DR 500 mg BID

One dose taken *x* hours after last dose	
15	Take dose and resume dosing
18	Take dose and resume dosing
21	Take dose and resume dosing
24, missed dose	Take make-up dose
24, two doses	Take doses and resume dosing
Two doses taken *x* hours after last dose	
27	Take doses and resume dosing
30	Take doses and resume dosing
33	Take doses and resume dosing
36	Take doses and resume dosing

DR 1000 mg BID

One dose taken *x* hours after last dose	
15	Take dose and resume dosing
18	Take dose and resume dosing
21	Take dose and resume dosing
24, missed dose	Take dose and resume dosing
24, two doses	Take doses and resume dosing
Two doses taken *x* hours after last dose	
27	Take doses and resume dosing
30	Take doses and resume dosing
33	Take doses and resume dosing
36	Take doses and resume dosing

REFERENCES

1. J. A. Cramer and B. Spilker (Eds.), *Patient Compliance in Medical Practice and Clinical Trials*. Raven, New York, 1991.

2. J. Urquhart, Role of patient compliance in clinical pharmacokinetics: a review of recent research. *Clin Pharmacokinet* **27**:202–215 (1994).

3. J. Urquhart, Variable patient compliance as a source of variability in drug response. *Excerpta Med Int Congress Series* **1178**:188–198 (1998).

4. J. Urquhart, The electronic event monitor. *Clin Pharmacokinet* **32**:345–356 (1997).

5. G. Levy, A pharmacokinetic perspective on medicament noncompliance. *Clin Pharmacol Ther* **54**:242–244 (1993).

6. R. Montgomery, Medication noncompliance: a deadly problem. *Official News Wire* (7/21/2005). `http://news.baou.com/main.php?action=recent&rid=20354` (accessed Jan. 30, 2006).

7. M. A. Kass, D. W. Meltzer, M. Gordon, D. Cooper, and J. Goldberg, Compliance with topical pilocarpine treatment. *Am J Ophthal* **101**:515–523 (1986).

8. H. Petri and J. Urquhart, Patient compliance with beta-blocker medication in general practice. *Pharmacoepid Drug Saf* **33**:251–256 (1994).

9. D. M. Waterhouse, K. A. Calzone, C. Mele, and D. E. Brenner, Adherence to oral tamoxifen: a comparison of patient self-report, pill counts, and microelectronic monitoring. *J Clin Oncol* **11**:189–197 (1993).

10. P. Rudd, S. Ahmed, V. Zachary, C. Barton, and D. Bonduelle, Improved compliance measures: applications in an ambulatory hypertensive drug trial. *Clin Pharmacol Ther* **48**:676–685 (1990).

11. B. Spilker, Methods of assessing and improving patient compliance in clinical trials, in *Patient Compliance in Medical Practice and Clinical Trials*, J. A. Cramer and B. Spilker B (Eds.). Raven Press, New York, 1991, pp. 37–56.

12. L. Gordis, Conceptual and methodologic problems in measuring patient compliance, in *Compliance in Health Care*, R. B. Haynes, D. W. Taylor, and D. L. Sackett (Eds.). Johns Hopkins University Press, Baltimore, 1979, pp. 23–45.

13. D. B. Christensen, Drug-taking compliance: a review and synthesis. *Health Serv Res* **13**:171–87 (1978).

14. U. E. Westfall, Methods for assessing compliance. *Top Clin Nurs* **7**:23–30 (1986).

15. R. E. Grymonpre, C. D. Didur, P. R. Montgomery, and D. S. Sitar, Pill count, selfreport, and pharmacy claims data to measure medication adherence in the elderly. *Ann Pharmacother* **32**:749–774 (1998).

16. K. D. Burney, K. Krishnan, M. T. Ruffin, D. Zhang, and D. E. Brenner, Adherence to single daily dose of aspirin in a chemoprevention trial. An evaluation of self-report and microelectronic monitoring. *Arch Fam Med* **5**:297–300 (1996).

17. B. J. Mason, J. R. Matsuyama, and S. G. Jue, Assessment of sulfonylurea adherence and metabolic control. *Diabetes Educ* **21**:52–57 (1995).

18. L. L-Y. Lim, Estimating compliance to study medication from serum drug levels: application to an AIDS clinical trial of zidovudine. *Biometrics* **48**:619–630 (1992).

19. J. Urquhart and E. De Klerk, Contending paradigms for the interpretation of data on patient compliance with therapeutic drug regimens. *Stat Med* **17**:251–267 (1998).

20. J. A. Cramer, R. H. Mattson, M. L. Prevey, R. D. Scheyer, and V. L. Ouellette, How often is medication taken as prescribed? A novel assessment technique. *JAMA* **261**:3273–3277 (1989).

21. W. Kruse, P. Koch-Gwinner, T. Nikolaus, P. Oster, G. Schlierf, and E. Weber, Measurement of drug compliance by continuous electronic monitoring: a pilot study in elderly patients discharged from hospital. *J Am Geriatr Soc* **40**:1151–1155 (1992).

22. J. R. Matsuyama, B. J. Mason, and S. G. Jue, Pharmacists' interventions using an electronic medication-event monitoring device's adherence data versus pill counts. *Ann Pharmacother* **27**:851–855 (1993).

23. I. Arnet and W. E. Haefeli, Overconsumption detected by electronic drug monitoring requires subtle interpretation. *Clin Pharmacol Ther* **67**:44–47 (2000).

24. D. Matsui, C. Hermann, J. Klein, M. Berkovitch, N. Olivieri, and G. Koren, Critical comparison of novel and existing methods of compliance assessment during a clinical trial of an oral iron chelator. *J Clin Pharmacol* **34**:944–949 (1994).

25. J.-M. R. Detry, P. Block, G. De Backer, J.-P. Degaute, and R. Six, Patient compliance and therapeutic coverage: amlodipine versus nifedipine (slow-release) in the treatment of angina pectoris. *J Int Med Res* **22**:278–286 (1994).

26. S. A. Eisen, D. K. Miller, R. S. Woodward, E. Spitznagel, and T. R. Przybeck, The effect of prescribed daily dose frequency on patient medication compliance. *Arch Intern Med* **150**:1881–1884 (1990).

27. W. Kruse and E. Weber, Dynamics of drug regimen compliance—its assessment by microprocessor-based monitoring. *Eur J Clin Pharmacol* **38**:561–565 (1990).

28. W. Kruse, W. Eggert-Kruse, J. Rampmaier, B. Runnebaum, and E. Weber, Dosage frequency and drug-compliance behaviour—a comparative study on compliance with a medication to be taken twice or four times daily. *Eur J Clin Pharmacol* **41**:589–592 (1999).

29. T. Mengden, B. Binswanger, T. Spuhler, B. Weisser, and W. Vetter, The use of selfmeasured blood pressure determinations in assessing dynamics of drug compliance in a study with amlodipine once a day, morning versus evening. *J Hypertens* **11**:1403–1411 (1993).

30. T. J. Steiner, T. Catarci, R. Hering, T. Whitmarsh, and E. G. M. Couturier, If migraine prophylaxis does not work, think about compliance. *Cephalalgia* **14**:463–464 (1994).

31. A. Ahmad, E. D. Boudinot, and W. E. Barr, R. C. Reed, and W. R. Garnett, The use of Monte Carlo simulations to study the effect of poor compliance on the steady state concentrations of valproic acid following administration of enteric-coated and extended release divalproex sodium formulations. *Biopharm Drug Disp* **26**:417–425 (2005).

32. H. Kastrissios and T. F. Blaschke, Medication compliance as a feature in drug development. *Annu Rev Pharmacol Toxicol* **37**:451–475 (1997).

33. W. Wang, Patient compliance and its impact on steady state pharmacokinetics, in *Applied Statistics in the Pharmaceutical Industry Using S-Plus*, S. Millard and A. Krause (Eds.). Springer-Verlag, New York, 2001, pp. 217–231.

34. P. Girard, T. F. Blaschke, H. Kastrissios, and L. B. Sheiner, A Markov mixed effect regression model for drug compliance. *Stat Med* **17**:2313–2333 (1998).

35. D. A. Hughes and T. Waley, Predicting "real world" effectiveness by integrating adherence with pharmacodynamic modeling. *Clin Pharmacol Ther* **74**:1–8 (2003).

36. E. N. Jonsson, J. R. Wade, G. Almqvist, and M. O. Karlsson, Discrimination between rival dosing histories. *Pharm Res* **14**:984–991 (1997).

37. J. Lu, J. M. Gries, D. Verotta, and L. B. Sheiner, Selecting reliable pharmacokinetic data for explanatory analyses of clinical trials in the presence of possible noncompliance. *J Pharmacokinet Pharmacodyn* **28**:343–362 (2001).

38. S. Mu and T. M. Ludden, Estimation of population pharmacokinetic parameters in the presence of non-compliance. *J Pharmacokinet Pharmacodyn* **30**:53–81 (2003).

39. D. Soy, S. L. Beal, and L. B. Sheiner, Population one-compartment pharmacokinetic analysis with missing dosage data. *Clin Pharmacol Ther* **76**:441–451 (2004).

40. L. B. Sheiner and D. B. Rubin, Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* **57**:6–15 (1995).

41. L. A. Kenna and L. B. Sheiner, Estimating treatment effect in the presence of noncompliance measured with error: precision and robustness of data analysis methods. *Stat Med* **23**:3561–3580 (2004).

42. W. Wang, F. Hsuan, and S. Chow, The impact of patient compliance on drug concentration profile in multiple dose. *Stat Med* **15**:659–669 (1996).

43. W. Wang and S. Ouyang, The formulation of the principle of superposition in the presence of non-compliance and its application in multiple dose pharmacokinetics. *J Pharmacokinet Pharmacodyn* **26**:457–469 (1998).

44. P. Girard, L. B. Sheiner, H. Kastrissios, and T. F. Blascke, Do we need full compliance data for population pharmacokinetic analysis? *J Pharmacokinet Pharmacodyn* **24**:265–282 (1996).

45. B. Vrijens, E. Tousset, R. Rode, R. Bertz, S. Mayer, and J. Urquhart, Successful projection of the time course of drug concentration in plasma during a 1-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *J Clin Pharmacol* **45**:461–467 (2005).

46. L. A. Kenna, L. Labbe, J. S. Barrett, and M. Pfister, Modeling and simulation of adherence: approaches and applications in therapeutics. *Am Assoc Pharm Sci J* **7**:E390–E267 (2005).

Graphical Displays for Modeling Population Data

E. NICLAS JONSSON, MATS O. KARLSSON, and PETER A. MILLIGAN

7.1 INTRODUCTION

Nonlinear mixed effects modeling of pharmacokinetic (PK) and pharmacodynamic (PD) data is often not a trivial activity. The hurdle of understanding the statistics and the general methodology involved in building models is one challenge; and learning the necessary software tools is another. Fortunately, many of these aspects are general (although complicated) in the sense that they are applicable to numerous analyses and data types, making these skills important for any pharmacometrician to acquire. On the other hand, each new analysis and each new data set is unique and it is as important to understand the structure of the data to be analyzed as it is to understand the methodology to be used in the analysis. This is one instance where graphics can play a crucial role. Often the pharmacometrician is faced with the situation where the strategy and direction of model building is contingent on the data available at the time of initiation (i.e., models "are fit" to data). If you have not seen a picture of the key aspects of your data, then you have not understood what is required of you as a pharmacometrician! The visualization of the data will guide and determine the model building strategy and will therefore directly influence the efficiency with which the final model is derived.

Graphics is also an important diagnostic tool in model development. Plots of residuals, predictions, and other variables will inform the pharmacometrician whether the model addresses all relevant aspects of the data or if some part(s) of the model needs further attention.

Once the final model has been derived and the analysis has achieved a satisfactory conclusion, the pharmacometrician has to face the important task of communicating the results of the analysis to nonpharmacometricians. This involves translating the mathematical relationships of the model into a form that directly addresses issues that are relevant, for example, responder rates in various patient subpopulations or a suggestion for the design of the next study in a drug

development program. The important thing to remember here is that the intended audience should not have to be distracted by mathematics and Greek symbols of the model while assessing the implications of the new findings. Graphics is a very powerful tool to accomplish this.

Throughout this chapter we use two example data sets. The first is a real data set from a PK study in 73 individuals with an average of ten observations per individual. Each individual was studied on one to seven occasions. The second example is a simulated data set with an ordered categorical response variable and is described in greater detail in Section 7.4.7.

The graphs and examples are geared toward NONMEM simply because NONMEM is the most widely used computer program for population PK/PD analysis. The principles, on the other hand, are quite general and should be easily adoptable for use with other software employing the same methodological strategy as NONMEM does.

For the continuous data example, all the graphs are exemplified using PK data. However, the graphs are also as suitable for PD data. When the term simulation is used it refers to stochastic simulations. Finally, we have limited the chapter to regular two-dimensional graphs and omitted three-dimensional or interactive graphing techniques. The reason is that multidimensional relationships can be handled quite successfully by multiple two-dimensional graphs, thereby becoming much easier to understand. Interactive graphics, on the other hand, can potentially be quite useful in the explorative phases of a data analysis project but do not lend themselves for presentation or reporting purposes.

Following this introduction is a general section presenting techniques for handling the special aspects of population type data. The remaining sections of the chapter are organized around three phases in the life span of any analysis: *before, during, and after*. The activities in these phases are distinctly different and require different types of graphical approaches. For each of these phases we discuss graphs we believe are almost always useful. Section 7.4.7 addresses the specifics of the graphical display of categorical data. Visualizing this type of data, both from an exploratory as well as from a diagnostic perspective, requires different techniques compared to continuous data and is underused in the area of nonlinear mixed effects modeling. At the end of the chapter are appendixes detailing some NONMEM code required to generate data subsequently presented in graphs shown in the main body text.

7.2 CHARACTERISTICS OF INFORMATIVE DISPLAYS OF POPULATION DATA

The data used in population PK/PD analysis has some important features that need to be taken into account when plotting. The data is hierarchical, meaning that the observations are grouped according to the individual from whom they originated. This is the fundamental reason why a nonlinear mixed effects modeling approach was selected as the method of analysis in the first place. Our graphical methods must recognize this fact.

Population PK/PD data is heterogeneous. For example, individuals can vary in their drug elimination capacity and observed responses compared to other indivi-

duals as well as within themselves between study occasions. Similarly, observations vary due to dosing history differences, assay variability, or sample handling differences. An informative graphical display needs to handle this variability without masking it.

Population PK/PD data is multidimensional. In an analysis of PK data, the most obvious predictor we have is time. In an analysis of PD data, we have time and drug exposure as the fundamental independent variables. What should not be forgotten, however, is that there may be other potential predictors that can explain the observed variability (e.g., body weight, sex, age, and other covariates), some of which also vary with time. Again, we must use graphical methods that can accommodate this situation.

One aspect of population data that is often overlooked from the point of view of graphical displays is that we may have a large amount of data points. Regardless of how cleverly we devise our graphs, they will be useless if the data points form a dense, uninformative "shotgun blast" between the axes.

We must also be prepared to handle different types of response variables. For example, it is becoming increasingly common to model categorical data within the framework of nonlinear mixed effects models. From a plotting point of view, this type of data is distinctly different from continuous data and we need to account for this fact.

Figure 7.1 shows a graph of observed concentrations versus time after dose (see Section 7.2.1 for further discussion of time scales). Each individual's data points
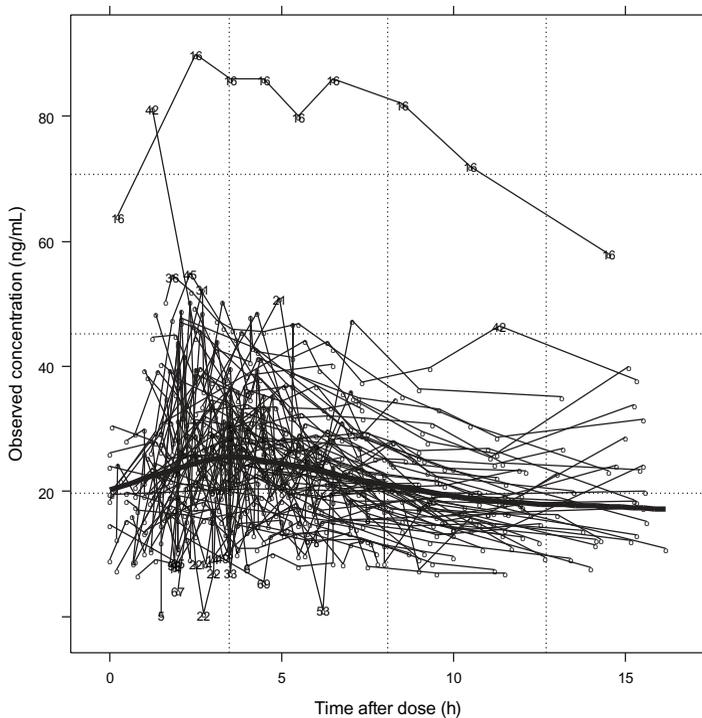
FIGURE 7.1   The plot illustrates some important aspects of the graphical display of population data. See the text for details.

are plotted using open circles (or ID numbers, which will be explained later) and connected with thin solid lines. The underlying trend in the data is visualized by a smooth nonparametric curve (the thick, black, solid line). The smooth line in this case is a loess regression line but other smooth curves, for example, running averages, will accomplish the same thing. (Sometimes it may be necessary to adjust the amount of smoothness of such trendlines—accomplished by changing the tuning parameters of the smooth curve, for example, bandwith or span. The rule of thumb is that the curve should be as smooth as possible without hiding important trends and an initial trial and error approach is usually sufficient to find an appropriate degree of smoothness.) Extreme data points are labeled with the corresponding individual's ID number. In this case, the extreme data points are defined based on the residuals from the fit of the smooth curve. Points outside the 2nd and 98th percentile of the residual distribution were defined as extreme. The thin, dashed grid lines are a visual reference grid, which is intended as a reference and not to help reading off values on the axes. This is useful if we have other plots we want to compare the current one to, but this requires the grid lines to be the same relative to the axis scales in the two graphs to be compared.

There is a lot of information contained within this graph. Individuals are explicitly recognized since their data points are connected. All individual data points are plotted, which makes it possible to appreciate the variability in the data without resorting to summary measures such as mean concentration curves ±95% confidence bands. We can also track the limits of the variability since the extreme individuals can be identified through their ID numbers.

Figure 7.1 is not very busy but it can easily be appreciated that graphical displays of this kind are at risk of becoming so. The use of open circles (or other nonsolid symbols) helps somewhat. Dashed lines may be perceived as a better choice than solid lines to reduce the amount of "ink" in the graph. However, dashed lines are harder to track among many other dashed lines and will therefore destroy the possibility to identify individuals as the data becomes denser, sooner than is the case with solid lines. The solution to the graphical display of a massive amount of data is actually not to plot all of it—so-called data dilution. The question therefore arises as how to omit data in an objective way. The obvious solution is to select a random subset of the individuals to display from the full set of individuals. This has been done in Figure 7.2, which displays a random selection of approximately 50% of the individuals in the original data set. It is quite clear that Figure 7.2 is giving a different picture from Figure 7.1. In particular, the most extreme individual (ID number 16) and the individual with the singly most extreme data point (ID number 42) are not included in this sample. This shows that displaying a completely random subset may actually hide the very information we are interested in. One could argue that omitting 50% of the individuals is too much, but then it must be remembered that it may be necessary to exclude an even larger fraction if the data set is large. An alternative approach is to use stratified randomization in which it is made sure that the random subset retains all extreme individuals and only omits individuals who do not seem to contain any unique information. The definition of extreme *individuals* is similar to the definition of extreme data points. It is based on the residuals from the fit of the smooth curve. Individuals who have all their residuals inside (in this case) the 1st and 99th percentile of the residual distribution were regarded as eligible for exclusion from the graphical display. This is shown in Figure 7.3. The number
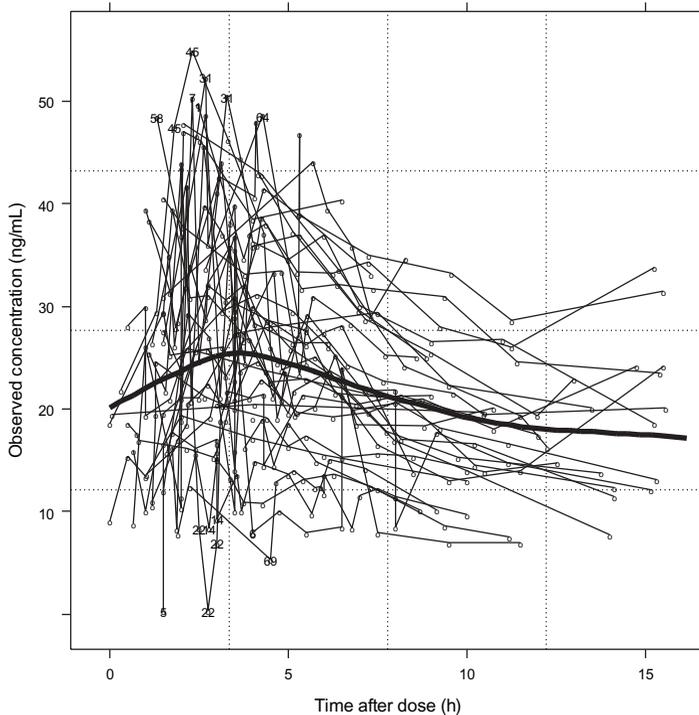
FIGURE 7.2 The plot illustrates data dilution based on a completely random subset selection.

of individuals in Figure 7.3 is approximately the same as the number in Figure 7.2. However, Figure 7.3 faithfully retains the same impression of variability as Figure 7.1 and the smooth curve, which is the same in all three figures (i.e., based on all data), provides the central tendency impression. To ensure that any data dilution does not distort or hide important information, it may be useful to create a handful of diluted graphs and make certain that an appropriate set of sampling parameters has been selected. This should, however, only be necessary at the outset of an analysis or if the data set is amended or changed during an analysis.

That leaves the multidimensionality aspect unaddressed. The basic problem is that we may have more than one potential predictor. One possibility is to use three-dimensional (3D) graphs, but they are difficult to interpret and only allow three dimensions. Another possibility, which turns out to work very nicely in this setting, is multipanel conditioning plots (as implemented in Trellis Graphics library in S-Plus (Insightful Corporation, Seattle, WA) and in the lattice library in R (www. r-project.org)). With this type of graph, one variable is plotted versus another *given* intervals of a third (or fourth, or fifth, etc.) variable. An example is shown in Figure 7.4.

The observed concentration is plotted versus time after dose given three ranges of creatinine clearance (low, medium, and high as defined by the intervals given by the 33rd and 67th percentile) as indicated by the *strips above each panel*. The visual reference grid makes it easy to compare the data in each of the panels. The plot
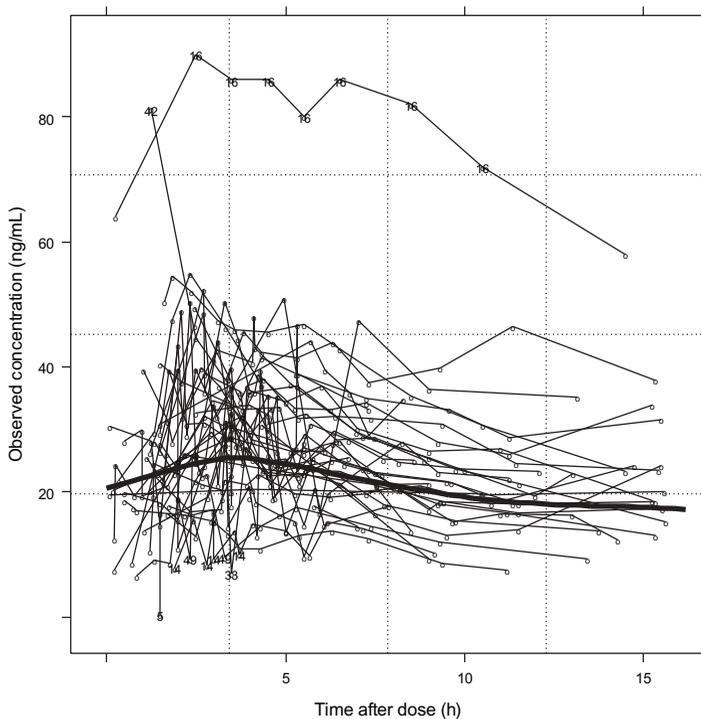
FIGURE 7.3 The plot illustrates data dilution based on a stratified, randomly selected subset.
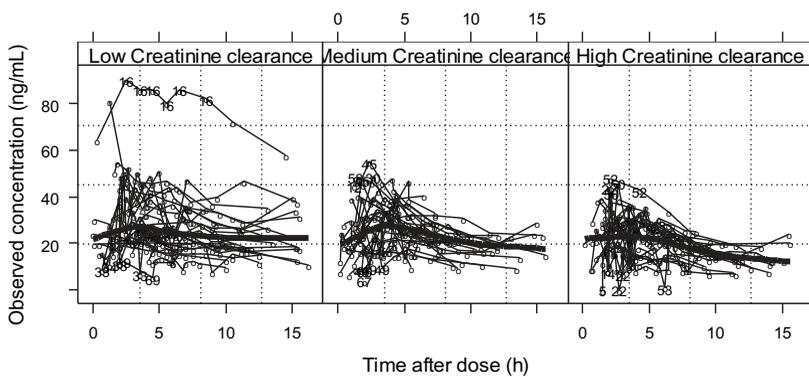
FIGURE 7.4 A multipanel conditioning plot of concentration versus time given ranges of creatinine clearance.

could easily have been expanded to use two conditioning variables, for example, creatinine clearance and age, in which case there would have been three panels for each age range, low, medium, and high creatinine clearance. For further information about the methodology of multipanel conditioning, please refer to the excellent book by Cleveland (1).

For the PK example used it is known that drug clearance is strongly related to the renal function and by comparing the slopes of the smooth curves in the three panels of Figure 7.4 we can certainly expect that this is to be true also in this data set. (In the High Creatinine clearance panel, the slope of the smooth curve seems to be steeper, i.e., a shorter half-life, which may indicate that the clearance is higher in these individuals.)

7.2.1   Time Scales

Time is the fundamental predictor in PK/PD models and therefore deserves some special attention. In NONMEM it is possible to specify the time points for observations and dosing events in the form of date and clock time. This is very convenient as this is the form in which the data is often stored in clinical databases. These dates and clock times are converted to decimal times in the preprocessing stages of the execution of a NONMEM run, and it is these decimal times that are used by NONMEM in the minimization procedure and that are provided in the tabulated output. From a plotting point of view, dates and clock times are not easy to work with. Except for cases with diurnal variations and/or annual rhythms (2, 3), the extra dimension offered by dates and clock times are unnecessary and may actually make it harder to visualize the data in an informative way.

Another aspect of time scales is that they are relative to something. The default behavior in NONMEM (after any time and date specifications have been converted to decimal times) is that the time scale in each individual is relative to that individual's first data record. In graphical displays, on the other hand, it may sometimes be more useful to use time after the last dose as the time scale, that is, to use the time of a specific dose (previous dose, last in a specific dosing period, etc.) as the reference point, since this creates a natural order in the time variable relative to predictions and residuals and hence more informative graphical displays.

7.3   BEFORE ANALYSIS

7.3.1   Tasks at Hand

Before any "real" analysis work can commence, there are at least two, potentially unrelated, issues that need to be resolved. The first is to make sure that the data file to be used in the analysis is correct. The second is that the pharmacometrician needs to explore the data to be analyzed. Both of these before-analysis tasks benefit from a graphical approach.

7.3.2   Data Checkout

One aspect of data checkout is finding errors in the data. Errors can occur either in the database(s) from which the data file is constructed or in the transcription from that database. The latter type of error can be of two types, either in individual values or in the structure of the data file as such. All three types of errors may pass unrecognized by NONMEM and additional data checkout is therefore necessary. The structure of the data set informs NONMEM where and when the doses and

dependent variables were observed, the values of independent predictor variables, and so on. Clearly this type of information will have a large impact on the results of the analysis and we need to make sure it is correctly specified.

Some errors in the structure of the data file, for example, positive drug concentration observations before the first dose, become apparent through error messages from the PREDPP library of NONMEM. Other types of mistakes are usually only found by inspecting the data file and knowing what the data should be, for example, the wrong covariate value(s) for an individual. Another way of finding mistakes is through association and plotting, meaning that by looking at a picture of the data we can see irregularities in some individuals that may indicate errors. The last approach is especially useful for the distribution of dosing and observation events.

One useful type of graph for finding errors in the structure data file is to plot each individual's ID number versus each column (variable) in the data file. Depending on the variable, the expected pattern is different. For example, when plotting the ID versus the AMT column (which is positive at time points when a dose is administered and zero otherwise), it is expected that each individual will have AMT values of 0 (observation events) and values for the administered doses, while when plotting covariates such as WT and AGE there should probably not be any values at zero. Figure 7.5 shows an example of this graph. Each individual has one data point for each of its rows in the data set. To avoid having all the identical data points from a single individual becoming superimposed, the data points have been jittered, meaning that for display purposes they have had an element of random noise added to them. The power of this type of graph lies in the association, an individual who is much different from the other individuals is usually obvious. However, it is difficult to find small errors, especially those that do not make the individual stand out compared to the others. One way to make the identification of extreme individuals even easier is to order the ID number of the *y*-axis according to their mean value of the *x*-variable. If there are more than, say, 100 individuals in the data set, it is necessary to split the graph over two or more pages.

A critical aspect of PK data is the dosing history and the relative placement of doses and observations in time, and it is important to check that there are no gross errors in this part of the data specification. Dosing histories can quite conveniently be displayed using *event history diagrams*. An example is shown in Figure 7.6 in which ID numbers are plotted versus the AMT column. Each individual's values are connected with a dashed line (which of course will be horizontal) and each event is indicated with a symbol—circles for nondosing events and vertical bars for dosing events.

A problem with Figure 7.6 is that the horizontal dimension is dominated by the observations at later times, making it hard to see what is going on at earlier time points. One solution to this may be to use a log scale for the time axis. Another is to split the graph and look at different time periods separately. The latter has been done in Figure 7.7, where the data coded as belonging to occasion one is plotted separately. Here we can see that most individuals had five doses recorded in the data set prior to the first observation while six individuals had observations recorded directly after the first dose event. (This may indicate that these six individuals lacked the dosing history and that the first dose had to be assumed to be a steady-state dose.) It is also obvious that, for most individuals, occasion one was defined to cover
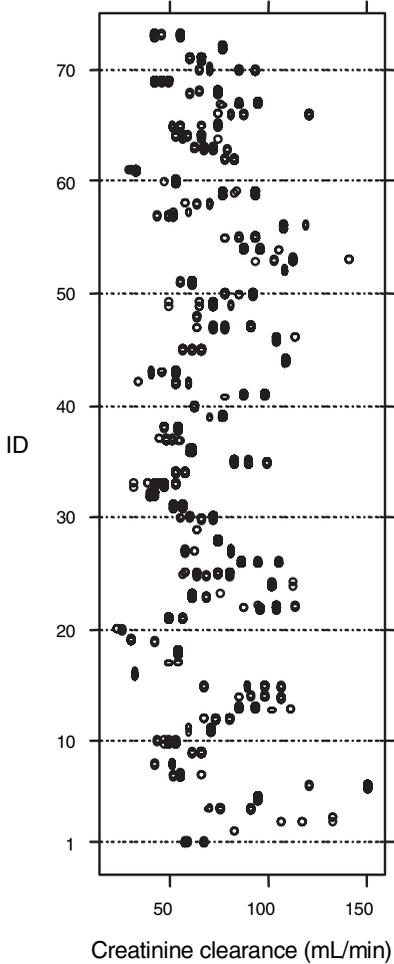
FIGURE 7.5   A data set checkout plot. The ID column in the data file is plotted versus the creatinine clearance column.
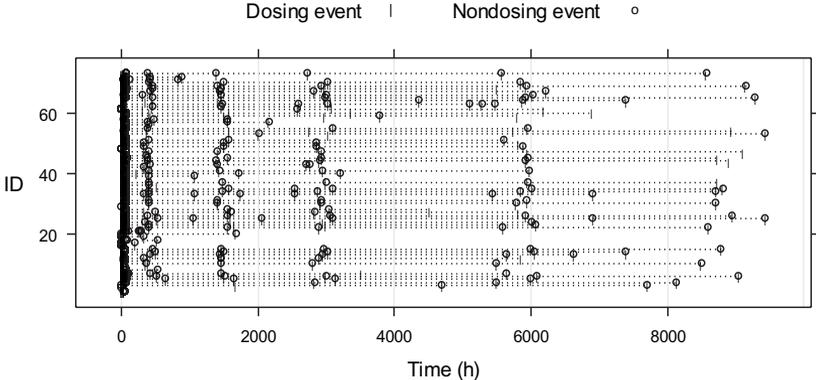

FIGURE 7.6   An event history diagram for all doses and observations in the data set.
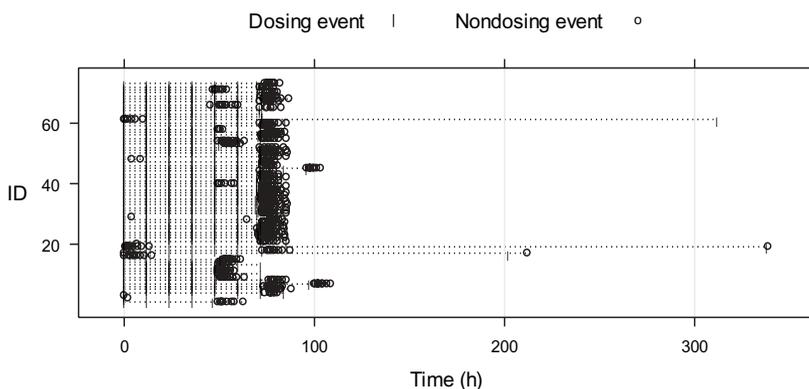
FIGURE 7.7   An event history diagram for the doses and observations coded in the data set as belonging to the first study occasion. See the text for details.

less than 100 hours while five individuals were different in this respect. Depending on how the other occasions were defined, this may have an impact on the quantification of interoccasion variability (4).

7.3.3   Data Exploration

Data exploration is a scientific exercise where we try to learn things about the data, for example, how the covariates are distributed and how they relate to each other. The exploratory data analysis also defines the population—and thereby the bounds for the validity of the model—and will form the basis for reporting the analysis to others. The data exploration is also important from an error finding point of view since some errors only become apparent when closely studying the data.

Exploratory data analysis in the field of population PK/PD analysis has received considerable attention in the literature over the years and the basic graphical tools for this exercise (e.g., histograms, scatterplot matrices, and QQ plots) have been described elsewhere (5–9) and will therefore not be detailed here.

One aspect that is particular to long-term outpatient clinical trials is time-varying covariates, and recently a framework for handling these has been presented (10). To visualize time-varying covariates, we need to use the original time scale, not time after dose. An example is shown in Figure 7.8, where each individual is plotted separately with a line that is horizontal until the value of creatinine clearance changes. (This is also the default way NONMEM handles time-varying covariates. It can be changed by including appropriate statements in a $BIND record in the NM-TRAN control stream.) Superimposed are the observed values and a smooth curve to visualize the underlying trend. In this case creatinine clearance appears to systematically increase with time, something to consider including in the model (10). However, care must be taken not to mistake this pattern for the case in which individuals with poor kidney function have a higher drop-out rate than individuals with relatively normal functioning kidneys. Care must also be taken not to let a skewed distribution of the covariate mislead the eye to see a pattern where none exists.
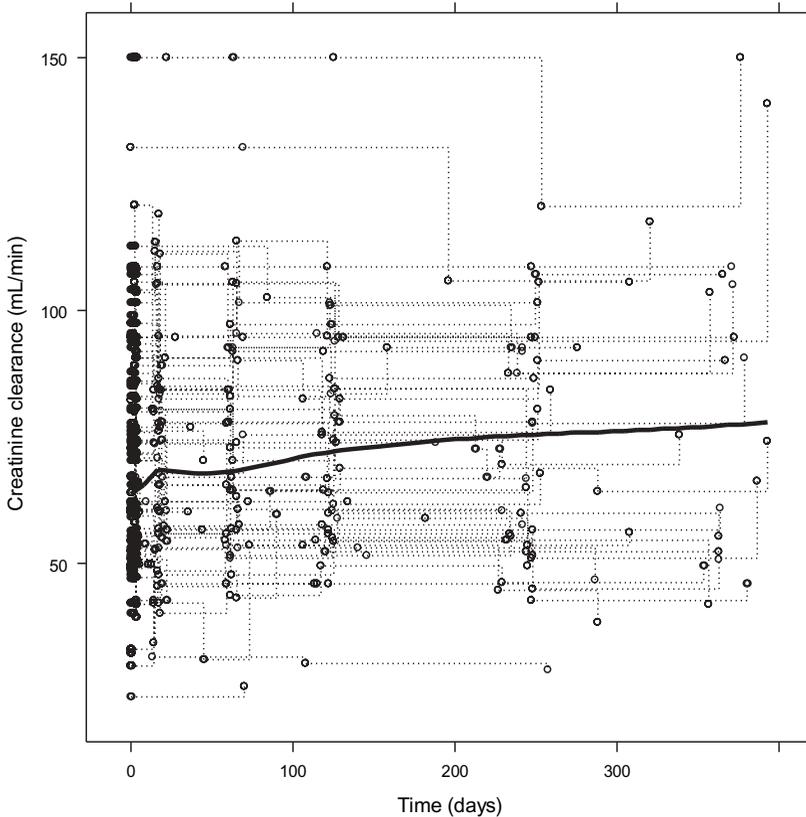
FIGURE 7.8   Creatinine clearance plotted versus time. The creatinine clearance values at each time point for each individual are plotted and connected with a dotted line. The solid black line is a smooth, nonparametric, regression line.

7.4   DURING ANALYSIS

Once the "real" analysis work starts, graphics should be an integrated part of the workflow. Graphics is used to suggest improvements to the model and is used to evaluate the benefits of the changes. There are other means of evaluating the importance of a change to the model, for example, statistical significance criteria, but only graphics can tell whether a model is appropriately describing the data.

During the model building phase, the pharmacometrician is constantly faced with two questions: (a) How can I improve the model to provide a meaningfully improved fit to the data? (b) Does the model violate any (statistical) assumptions, making it inappropriate? The first question is scientific and will depend on the application. For example, would the model benefit from a component that introduces time dependency in response? The second question is more general and applies to all analyses. One example is to choose an appropriate model for the residual variability (weighting). Karlsson and co-workers (9) compiled a comprehensive list of the assumptions involved in a NONMEM analysis together with suggestions

for how they could be checked. Many of the suggested approaches are based on graphical methods.

It is clear that most graphs in this phase of an analysis are only meant for the eyes of the pharmacometrician(s) and the questions answered by the graphs are quite technical. This means that the graphs can and perhaps need to be quite complicated and hard to understand for the uninitiated viewer.

7.4.1   Basic Plotting Variables

In the following we describe a number of graphs, grouped based on the question they are designed to answer. Before continuing, however, it is necessary to provide background and definitions for some of the important plotting variables:

$$\bar{P}_i = g(z_i) \tag{7.1}$$

$$P_i = q(\bar{P}_i, \eta_i) \tag{7.2}$$

$$\hat{\bar{y}}_{ij} = f(\bar{P}_i, x_{ij}) \tag{7.3}$$

$$\hat{y}_{ij} = f(P_i, x_{ij}) \tag{7.4}$$

$$y_{ij} = p(\hat{\bar{y}}_{ij}, \varepsilon_{ij}) \tag{7.5}$$

$$y_{ij} = p(\hat{y}_{ij}, \varepsilon^i_{ij}) \tag{7.6}$$

Typical individual parameters (Eq. (7.1)), $\bar{P}_i$, can be expressed as a function, $g(\ )$, of covariates, z_i. Note that there will be as many unique typical individual parameter estimates as there are unique combinations of z_i. If the model does not include any covariate relation, then $g(\ )$ will be a constant value.

Individual estimates of the parameters (Eq. (7.2)), P_i, can be expressed as a function $q(\ )$ based on the typical individual parameter estimate and the zero mean, symmetrically distributed variable η_i, whose standard deviation is ω. The values of η_i are obtained as posterior Bayes estimates conditioned on a set of typical individual parameter estimates, ω_s, and the data for the individual. The important property of the individual parameter estimates is that they will be shrunken toward the typical individual parameter estimate, the degree of shrinkage determined by the amount of information in the data from the individual relative to the size of ω. In the extreme case with no data at all, the individual estimate will be exactly the same as the typical individual estimate. In the other extreme, with an infinite amount of information, the individual parameter estimate will be independent of the typical individual estimates and ω.

The predictions based on the typical individual parameter estimates (denoted $\hat{\bar{y}}_{ij}$ in Eq. (7.3)) are, in NONMEM, called PRED. The size of the prediction depends on the individual's value of any independent variable (e.g., time) at observation j, x_{ij}. The individual predictions ($\hat{y}_{ij}$ in Eq. (7.4)), usually called IPRED, are not directly available in NONMEM but can easily be derived by the user in the NONMEM control stream. How this can be performed is described in Appendix 7.1.

The differences between the predictions and the observations are the residuals and the values of these will depend on the function $p(\ )$ and if the population or individual predictions were used to derive them. In Eqs. (7.5) and (7.6) they are

denoted ε_{ij} and ε_{ij}^{i}, respectively. The standard deviation of ε_{ij} is σ. The weighted residuals, WRES, obtained from NONMEM are based on the typical individual estimates of the parameters and are scaled such that they ought to have a mean of zero and unit standard deviation (*NONMEM User's Guide II* (11)). The individually weighted residuals, IWRES, are not directly available in NONMEM but can, similar to the IPREDs, easily be derived in the NONMEM control stream (see Appendix 7.1).

7.4.2   Overall Goodness of Fit Plots

Before progressing into the world of residuals, it is usually a good idea to get an overall impression of the model performance. A useful display for this purpose is shown in Figure 7.9. The observed data and the individual and population predictions are plotted versus time after dose. The data was diluted to about 25% using stratified randomization and the extreme data points in the left panel and the corresponding predictions in the middle and right panels are labeled with the ID number. If the two right panels are satisfactory, meaning that they look similar in some sense to the left panel, we know that the model is at least improving. For more detailed structural model diagnostics we have to resort to other plots (see below). We can also learn about the scope for improvement by adding covariate relations to the model. The individual predictions usually define the limits for what is achievable by adding covariates. Or phrased differently, we can rarely hope to bring the population predictions closer to the observations than the individual predictions are by the addition of covariates to the model, meaning that the difference between middle and right panels defines the scope for model improvement by covariates. A final benefit of this graph is that we can show it to nonpharmacometricians and expect them to understand what is going on, despite the fact that we present it as a diagnostic plot.
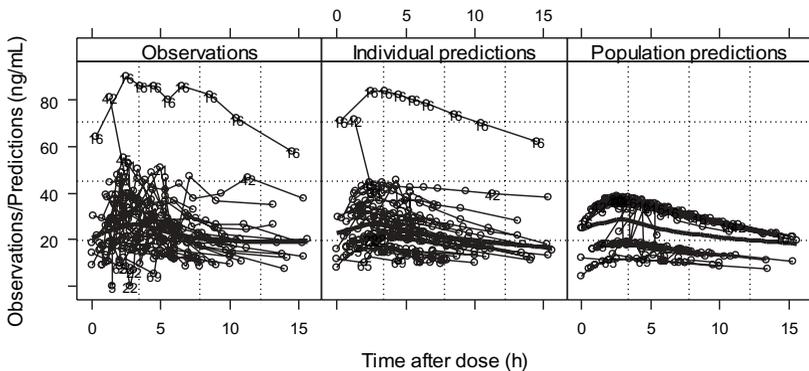

FIGURE 7.9   The figure shows the observed data and the individual and population predictions plotted versus time after dose. Included is approximately 25% of the total number of individuals (selected using stratified randomization). The thick solid line is a smooth curve based on all individuals. Data points that were judged extreme (see above) in the left panel are labeled with the ID number. The corresponding predictions are also labeled. The zigzag pattern most obvious in the right panel is due to the fact that some individuals received different doses at different study occasions.

Figure 7.9 is good for obtaining a general impression of the performance of the model. If we want to get a more detailed look at how the predictions match the observations, we can plot them against each other (Figure 7.10). The graph displays all data (no dilution). The line of identity (the solid bold line) is included as reference. Since the model does not contain any covariates, the differences in population predictions are only due to differences in dose amounts (left panel). When the data is taken into account, the predictions cover the same range as the observations (right panel). What we expect from this type of graph, if the model is fitting the data well, is that the data points are scattered evenly around the line of identity; that is, the average prediction goes through the middle of the observed data and the predictions and observations are near the line of identity. If this is the case, then the scatter around the line of identity in the left panel can basically only be reduced by the addition of covariates. The scatter in the right panel is also influenced by how we model the interindividual variability. An alternative if the plot is dominated by a few extreme data points is to use a log–log scale.

Note the placement of the axes: the observations on the y-axis and predictions on the x-axis. This is in line with Eqs. (7.5) and (7.6), where the y-variable is the
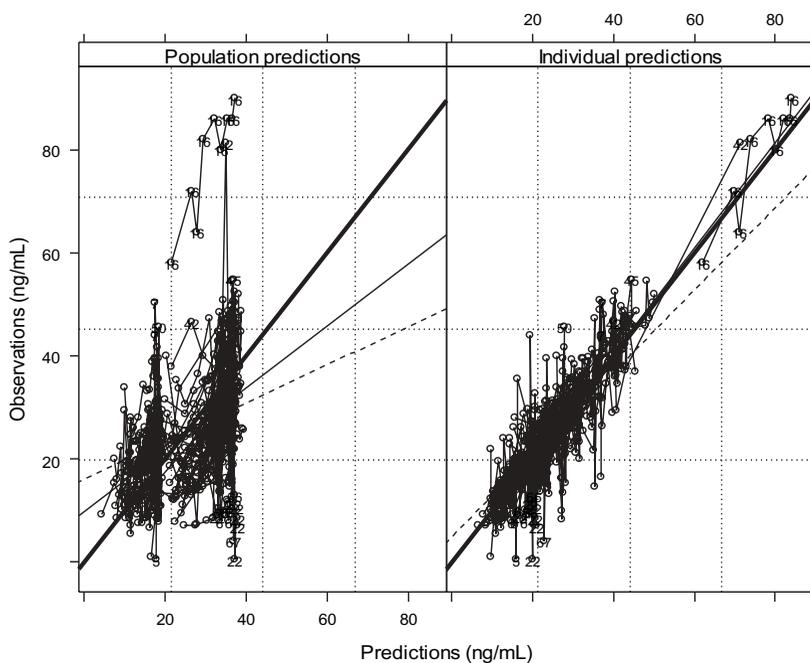
FIGURE 7.10   The observations are plotted versus the population and individual predictions, respectively. All data points are included in this graph (no dilution). The thick solid line is the line of identity. Data points that were judged extreme (see above) in the left panel are labeled with the ID number in both panels. The thin solid line is the linear regression curve obtained when the observations were regarded as the y-variable (the correct way). The dashed line is the linear regression curve obtained when the observations were regarded as the x-variable and the predictions as the y-variable (the wrong way).

observations and the unexplained variability is additive to the predictions (in the y-axis direction). Obviously the axis placement does not influence the data plotted, except that the picture will be turned around. However, it does matter which variable is regarded as the y-variable if we wish to add regression lines (linear or smooth curves). Such regression models assume that the unexplained variability is additive in the vertical direction. The thin solid, diagonal lines in Figure 7.10 are the regression lines obtained by regarding the observations as the y-variable (the correct way). The dashed, diagonal lines, on the other hand, were the results of regarding the observations as the x-variable (the wrong way). Clearly it makes a difference. Note that a deviation between a regression line and line of identity does not always indicate model misspecification. It may be a consequence of, for example, adaptive study designs or exponential parameter or error distributions (see below about simulations for inspecting diagnostic plot behavior).

To get a more detailed impression of the differences between the predictions and the observations and how these differences are distributed over the independent variable, we can plot the IWRES versus time after dose (Figure 7.11). With this graph, the choice of the independent variable is important. Had we chosen the original time scale (relative to the first data record in each individual), it would have been hard to interpret any trends that the graph shows but would, on the other hand, have allowed for the detection of time dependencies. An alternative,
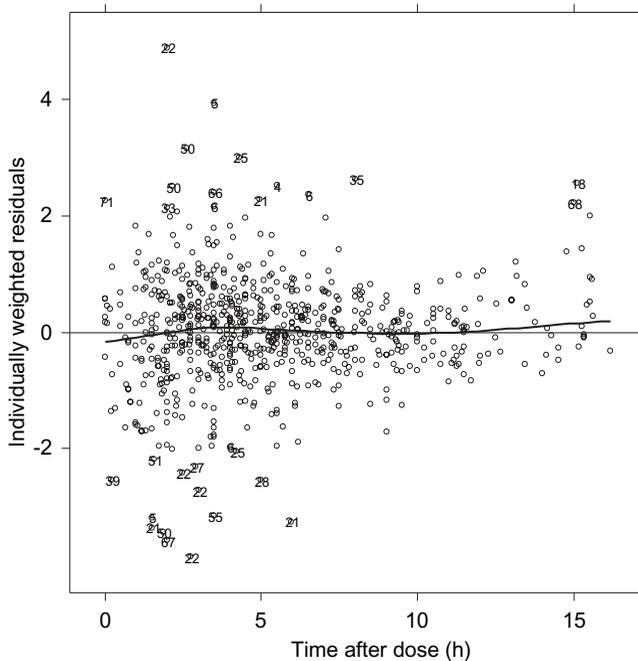

FIGURE 7.11   The individually weighted residuals versus time after dose. The solid thick line is a smooth nonparametric regression line. The horizontal thin line is the zero reference line.

especially in the case when the data per individual is sparse, is to plot the WRES instead of the IWRES. However, if any of the conditional estimation methods are used in NONMEM, WRES may be misleading. This is because they are always computed under the same assumption as the first-order (FO) estimation method is using. For a particular problem, if there is a difference in the parameter estimates between using the FO and a conditional estimation method, WRES should be used with caution to detect imperfections of models, as even appropriate models can display patterns indicating misspecification. As this is likely to happen mainly when individual data is relatively rich, IWRES is probably a better diagnostic. If we use the FO method, then the WRES are the residuals we should look at; if we use other methods, then they may not be appropriate to use. (For further details on the various estimation methods in NONMEM, please refer to *NONMEM Users Guide VII* (11)).

If the model is appropriate for the data, then Figure 7.11 should show no trends: that is, the data points should be evenly scattered around the horizontal zero-line, and the smooth curve should be approximately horizontal.

7.4.3 Residual Model Diagnostic Plots

In all types of data analysis there are assumptions made. In a parametric approach, like the one in NONMEM, many assumptions concern the handling of the residual error (9, 12) and, in a sense, the validity of the whole analysis rests on the degree to which we have accounted for the residual variability appropriately. The two most important assumptions in this respect are (a) that the residual variability is homoscedastic and (b) that the residuals are symmetrically distributed.

The assumption of homoscedasticity means that the residual variability should be constant over all available data dimensions (predictions, covariates, time, etc). If we observe heteroscedasticity, then we need to change the residual error model to account for this. In practice, this means that we should weight the data differently by using a different model for the residual variability.

Figure 7.12 is a useful graph for detecting problems with the residual variability model. The graph shows the absolute values of the individually weighted residuals versus the individual predictions. The use of the absolute values of the residuals assumes that there is a balance between positive and negative residual: that is, an appropriate structural model has been used. The smooth curve indicates if the underlying trend is different from horizontal. If the smooth curve has a pronounced positive slope, then the model should allow higher predictions to have a larger variability, for example, going from an additive model to a slope-intercept or a constant CV residual error model. If the slope is pronounced negative, then the residual error model should instead be more "additive," for example, going to a slope-intercept model from a constant CV residual error model. There is, of course, more to say about residual error modeling but this is beyond the scope of this chapter. Useful references in this respect are Refs. 9 and 12–14.

The second assumption concerns the distribution of the residual errors. Ideally, the weighted residuals should be normally, or at least symmetrically, distributed. Effective graphical displays to check this include histograms and quantile–quantile (QQ) plots (5). If a marked skewness is observed, it may indicate that a transformation, for example, a log transformation, of the data may be necessary (13).
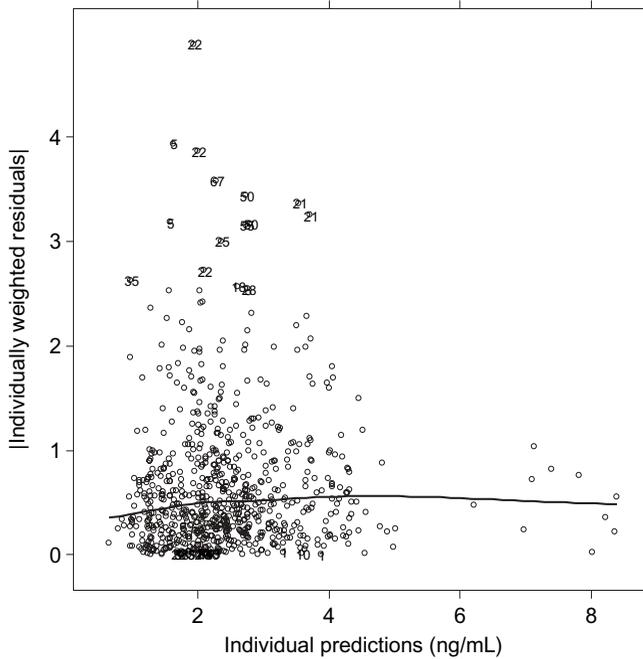
FIGURE 7.12   The absolute values of the individually weighted residuals versus the individual predictions. The solid line is a smooth, nonparametric regression line.

7.4.4 Interindividual Variability Model

Model building diagnostics for interindividual variability rely on investigation of η_i (or P_i). As η values can display considerable bias (shrinkage) when information about a parameter is lacking in an individual's data, basing decisions on these parameters can be misleading. Whenever shrinkage occurs, the individual parameter estimates tend to support the models used to generate them. One way of assessing the degree of shrinkage is to compare the ω estimate from the model with the SD of the η estimates. If the two values are similar, shrinkage is not a problem. However, if the individual estimates display a considerably lower variability, diagnostics based on η values should be treated with caution (and so should diagnostics based on IWRES). However, due to study design, sometimes only a subset of the studied subjects may contain substantial information about a particular parameter, and diagnostics may then be limited to that part of the data. The precision (SE) in individual η estimates is not provided as standard output in NONMEM but can be obtained subsequent to the fit (Appendix 7.2). Values of η with a large SE relative to ω are subject to more shrinkage than η values with small SEs. Note also that individual η values can represent a solution at a local minimum, a situation not easily diagnosed as there is no user influence over the initial estimates that are used in that estimation.

   Provided η values are deemed reliable, they can be used in scatter matrix plots to investigate correlations between parameters or in QQ plots or histograms to assess appropriateness of the selected parametric shape of parameter variability. Models

that include both interindividual and interoccasion variability offer a particular obstacle when it comes to use η values as diagnostic tools. If the number of occasions is large and the information per occasion high, η values for interindividual variability can be diagnosed as described. For situations (or individuals) where the number of occasions is low and/or information per occasion is scarce, diagnosis based on individual-specific and occasion-specific parameter estimates should proceed with caution.

7.4.5 Covariate Model

Incorporating covariates in a population PK/PD model is often an important part of the model building process and is often also an overall aim of the analysis.

Graphical displays of parameters versus covariates are appealing since we usually have an intuitive understanding of the plotted variables. Depending on whether the covariate is continuous or categorical, we need to use different plot types. Figure 7.13 is a plot of the unexplained variability in CL, expressed as the η for CL, versus creatinine clearance and sex. Creatinine clearance is a continuous variable and can be displayed using a typical bivariate xy plot, while sex is categorical and is displayed using a box and whiskers plot. The latter plot type is useful for categorical data. The solid symbol in the center of the box shows the median value. The box itself is limited by the interquartile range (25th to 75th percentile). The whiskers, the dashed lines going up and down from the box, extend 1.5 interquartile ranges from the box *or* to the most extreme data point. If there are any data points beyond the whiskers, these are plotted individually. Since the amount of data in each category is crucial in the judgment of any differences between the categories, the width of the boxes in Figure 7.13 are set to be proportional to the corresponding number of individuals (there are considerably less females than males in the data set).

It is generally hard to let graphs like the ones in Figure 7.13 guide the covariate model building process. The problem lies in the fact that covariates tend to be correlated. This is to some extent illustrated in Figure 7.13. The parameter values
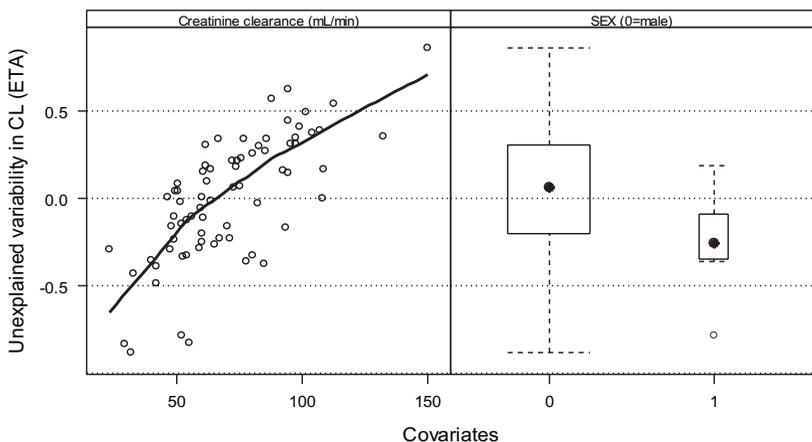

FIGURE 7.13 The unexplained variability in CL, expressed as η, when no covariates were included in the model, plotted versus creatinine clearance and sex. See the text for details.

come from a model that does not include any covariate relations. There seems to be a clear correlation between clearance and creatinine clearance but there also seems to be a difference between males and females. Since it can be expected that creatinine clearance and sex are correlated (we should have explored this in the before-analysis phase), the two apparent relations can be caused by the same mechanism. On the other hand, there *may* be two separate mechanisms involved. A way to sort this out is to include one of the covariate relations in the model, run it, and then construct the same graph again.

Figure 7.14 is based on a model in which creatinine clearance was included. The axis limits are the same as in Figure 7.13 and it is clear that the unexplained variability has decreased. At the same time it appears as if the sex relation is not as important anymore. On the other hand, had sex been included in the model instead of creatinine clearance then the picture would perhaps have looked the same. Again, this is the problem with using graphs to guide covariate model building. Clearly some other techniques are necessary (see other chapters in this book).

The best use of graphical displays of the above type is actually not to identify relations but rather to explore the shape of the relations, for example, linear or nonlinear, or to disprove them (6).

In Figures 7.13 and 7.14, the unexplained variability in clearance was expressed as the corresponding η value (obtained by posterior Bayes estimation). There are other alternatives, as shown in Figure 7.15.

Plotted are the individual estimates of clearance, the difference between the individual estimates of CL and the typical individual estimate of clearance, and the η for clearance. Without any covariates in the model it does not matter much which variable is used. With covariates, on the other hand, we should not use the individual estimates of the parameter. Equation (7.7) explains why.

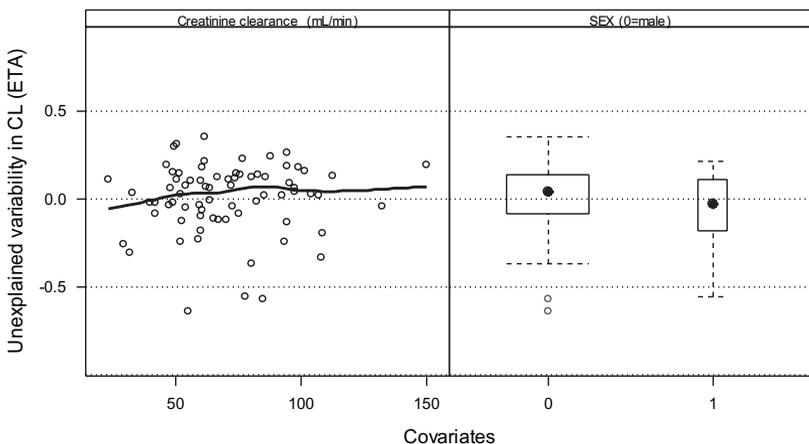$$P_i = \theta_P(1 + \theta_{\text{cov}}(x_i - \text{median}(x))) + \eta_i \tag{7.7}$$

FIGURE 7.14   The unexplained variability in CL, expressed as η, when creatinine clearance was included in the model, plotted versus creatinine clearance and sex.
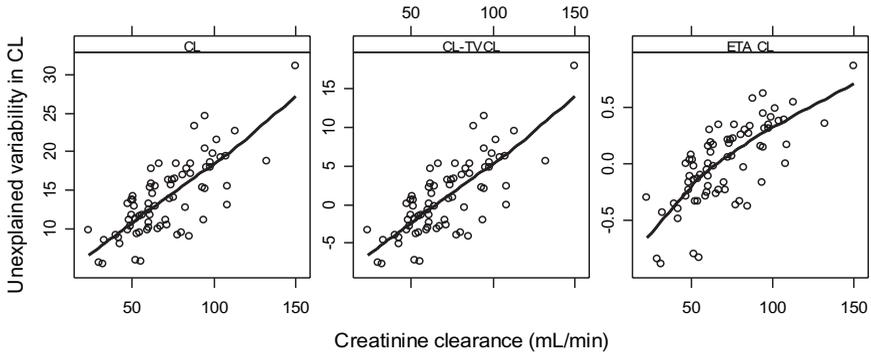
FIGURE 7.15   The variability in clearance without any covariates in the model plotted versus creatinine clearance. The three panels include different measures of the variability. See the text for details.
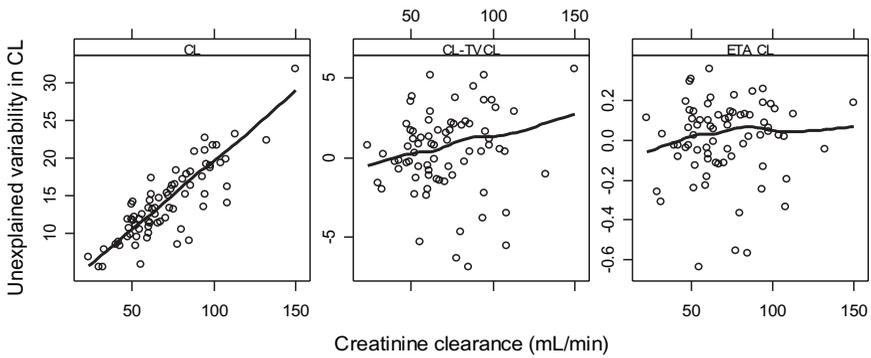

FIGURE 7.16   The variability in clearance in the model plotted versus creatinine clearance. The three panels include different measures of the variability. See the text for details.

θ_P is the typical value of the parameter P (defined as an individual having the median value of the covariate x), θ_{cov} is the coefficient for the covariate relation, and x_i is the ith individual's value of the covariate x. Clearly, if we plot the individual parameter estimates (P_i), obtained under the model in Eq. (7.7), versus the covariate x, we will see a trend. This is exemplified in Figure 7.16.

The individual parameter estimates in Figure 7.16 were obtained from a model similar to Eq. (7.7) and we can see that the individual estimates of clearance show a clear relation to creatinine clearance while the other two measures of unexplained variability are reduced in comparison. To summarize, once the model includes covariates, we should not plot the individual estimates versus covariates but rather something like the η values if the reason for creating the graph is to visualize potential relations between the unexplained variability and covariates.

Another point with respect to graphical display of covariate relations concerns interactions between covariates, that is, when the relation between the parameter and covariate depends on the value of another covariate. For example, if males and females have different relations between clearance and body weight. Given that the data set contains enough individuals to support the identification of interaction

effects, these can be visualized quite efficiently by the use of multipanel conditioning, as explained above.

7.4.6 Simulations as a Tool to Make Sense of Goodness of Fit Graphs

Diagnostic plots can only tell us about the deficiencies of the model and not about its adequacy unless we compare it to the same type of plots based on other contending models. In practice, if we cannot see any problems in our battery of diagnostic plots, then we assume the model is without substantive error. Therefore, it is important to use a multitude of graphs to inspect as many aspects as possible of the model. However, we need to be careful because a diagnostic plot may appear suboptimal even if the model is adequate. There are basically two reasons for this phenomenon.

The variable we are plotting, for example, the residuals, is not a good measure of the model adequacy: the reason usually being that there are approximations involved in the derivation of the variable, for example, as previously discussed for IWRES and WRES. Other issues with residuals are presented by Cox et al. (15) in the application of survival models in a nonlinear mixed effects environment.

In the judgment of diagnostic plots we rely implicitly on a notion of what the graph should look like if the model is adequate. For example, in a graph of the observations versus the predictions, we expect the data points to line up nicely around the line of identity; that is, the line of identity is our reference. The extent to which this is true depends on the type of observations, study design, and size of variability we are dealing with and the estimation method we use. If the estimation method involves approximations that lead to "nonstandard" graphs, then, clearly, we cannot improve the appearance of the graphs by changing the model (e.g., a plot of the population predictions versus the observations if there are no covariates available).

How do we know we are in this type of situation? One indication is that the aspect of the graph we are looking at does not change regardless of what we do with the model. The solution is strikingly simple: Use the model under consideration to simulate a new data set, analyze the simulated data set, and produce the same diagnostic plot as was done for the observed data (9). This plot will define a reference for the real data plot since we know that the plot of the simulated data was derived using the correct model. Examples are given in Figures 7.17 and 7.18. In Figure 7.17, the observations are plotted versus the predictions obtained from the model in which creatinine clearance was included as a linear predictor of clearance. In Figure 7.18, the predictions came from the same model applied to data simulated from the model in Figure 7.17.

There is a clear resemblance between the two figures although the one based on real data appears to be more variable. This may indicate that there are more covariate effects to be included in the model (based on the left panels) or that interoccasion variability would improve the model (based on the right panels) (4).

Quite often, it is enough to simulate just one realization of the data but sometimes, particularly if the data set is small and the variability is large, one realization is not enough to form a firm opinion about the adequacy of the model. One possibility then is to simulate many data sets (100–1000) and use them to construct prediction intervals to be superposed on the real data plot.
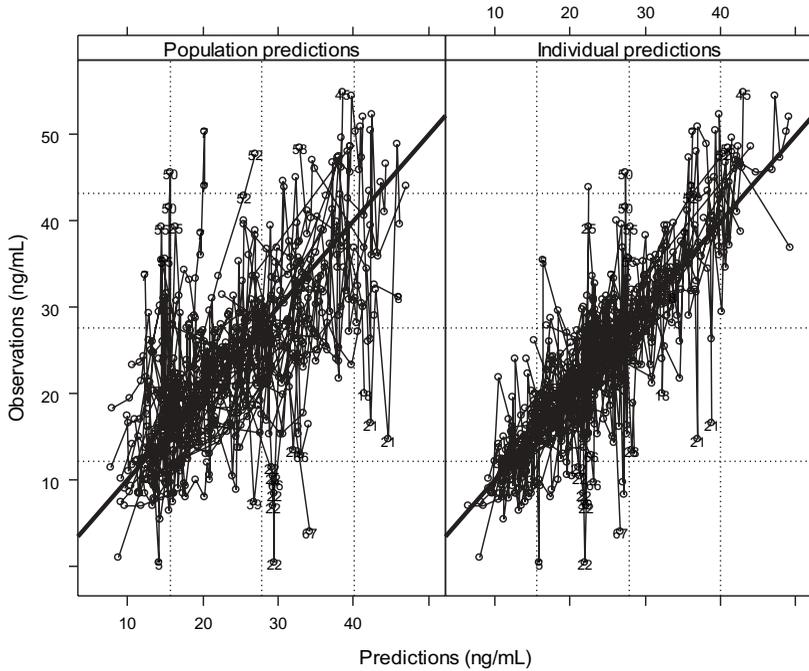
FIGURE 7.17 Plotted are the observations versus the predictions obtained with a model that includes creatinine clearance as a linear predictor of clearance. Since individuals 16 and 42 dominated the graph, they were excluded from this display.
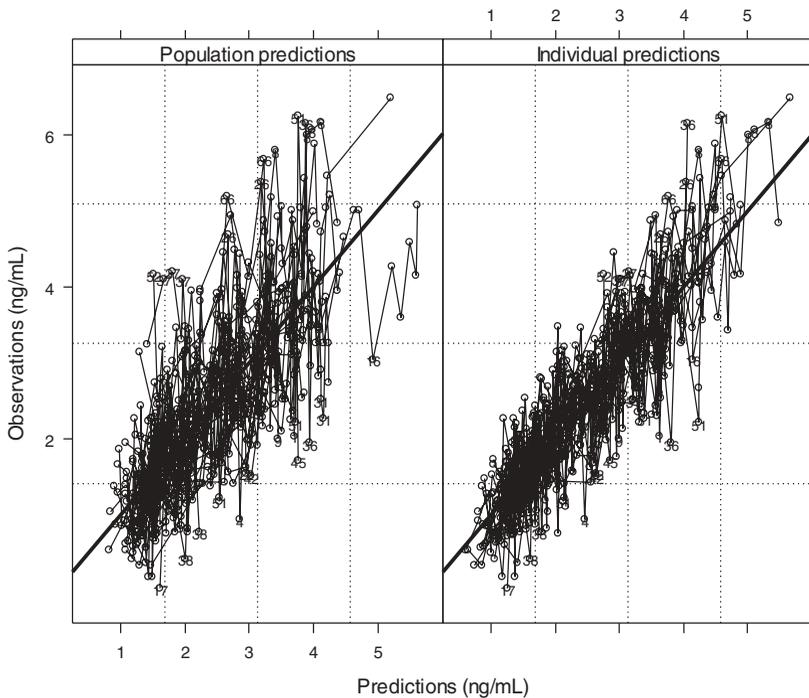

FIGURE 7.18 Plotted are the observations versus the predictions obtained with a model that includes creatinine clearance as a linear predictor of clearance applied to data simulated from the model used in Figure 7.17.

7.4.7   Categorical Data Poses Different Challenges

Categorical data is becoming increasingly common in population PK/PD analysis, especially ordered categorical data. Examples of such data are adverse events and efficacy measurements such as pain scales (16) or sedation scores (17). This section focuses on graphical methods for categorical type data.

When dealing with ordered categorical data, it is important to remember that an observation is regarded as being a realization of an underlying set of probabilities, with one probability existing for each level in the response variable. The model is describing these probabilities, and how they depend on the predictors, *and not the actual observations*. This will have consequences for the graphical methods we need to use. Basically, there is no point in plotting the observed values since the actual observations are the frequency (or probability) with which they are observed. In other words, we need to concentrate our plotting efforts on the observed and predicted probabilities.

In this section we use a different example data set. It consists of 1600 simulated categorical observations from 580 individuals. The response variable has six possible values (0–5). The data were simulated using an ordered logistical model as described by Zingmark et al. (17). The administered dose was the main predictor and could be 0, 25, 50, or 100 mg. The data were simulated with the dose being related to the outcome according to an E_{max} model (on the logit scale). In some of the graphs, the "wrong" model was used, meaning a linear model in dose rather than an E_{max} model.

7.4.7.1   Raw Data Visualization

For a quick look at the observations, without taking any predictors into account, we can use a regular histogram (Figure 7.19). This graph provides some information about the frequency of the observed data; for example, it is clear that category 5 is quite uncommon.

Since dose is a candidate predictor for these data, it is natural to take it into account when plotting (similar to time with PK data). One possibility is to use
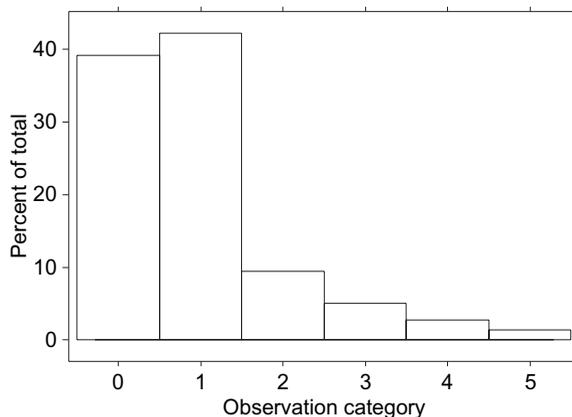

FIGURE 7.19   Histogram of the observed data. Each bar represents the frequency of the observed category.
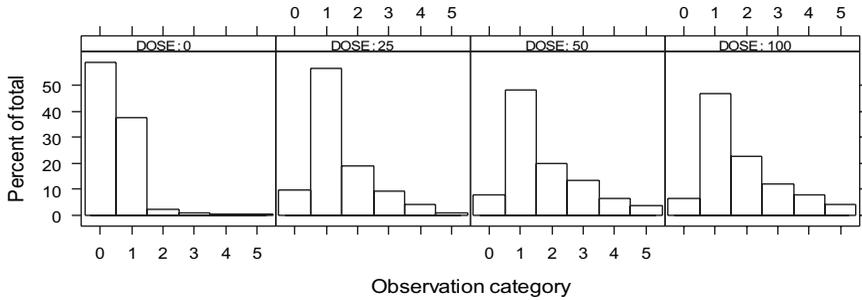
FIGURE 7.20   Histogram of the observed data given the dose level. Each panel shows a histogram of the data in one dose group.

multipanel conditioning, as shown in Figure 7.20. Each panel shows a histogram for one dose level. There seems to be a tendency of the "mass" of the distributions to shift to higher scores when the dose increases. If there were other potential predictors, similar graphs could have been produced for them as well, using intervals of any continuous predictors as the conditioning variable. Another alternative when there are multiple predictors is to construct multilevel conditioning plots, for example, plotting one histogram for each combination of dose and age interval. This can be rather involved and voluminous. Stacked bar charts of the observed probabilities of each score provide a more compact alternative to Figure 7.20.

Figures 7.20 and 7.21 display the same information but in different ways. The main benefit of the latter is it compactness. The possibility to see patterns, on the other hand, is hampered by the fact the there are no fixed reference points (except for the categories with the lowest number). This can be remedied by adding lines that connect the cumulative probabilities of each category (Figure 7.22). Apart from making it easier to see trends, this is also informative from the perspective of the model. A logistic model for ordered categorical data is usually defined in terms of the cumulative probabilities (score $\leq n$), that is, exactly what this graph is visualizing.

7.4.7.2   Assessing Goodness of Fit

To obtain predictions of the observed probabilities, we can simply use the proposed model to simulate a new data set at least as large as the observed one and compute the predicted probabilities in the same manner we computed the observed probabilities. NONMEM provides some functionality in this respect (11), but we find this simulation approach more straightforward to use. Figure 7.23 shows an example of this approach. Displayed are two stacked bar charts, one for the observed probabilities and one for the predicted probabilities. The predictions were obtained using a model in which dose was included as a linear function in the logit, that is, the wrong model. That the model is inappropriate is quite obvious.

With real data the pattern may not be as clear as in Figure 7.23, and we may wonder if the simulated data (the "predictions") is a fair description of the model. After all, it is only one random realization of the model. This can easily be checked by simulating more than one data set and computing the predicted probabilities
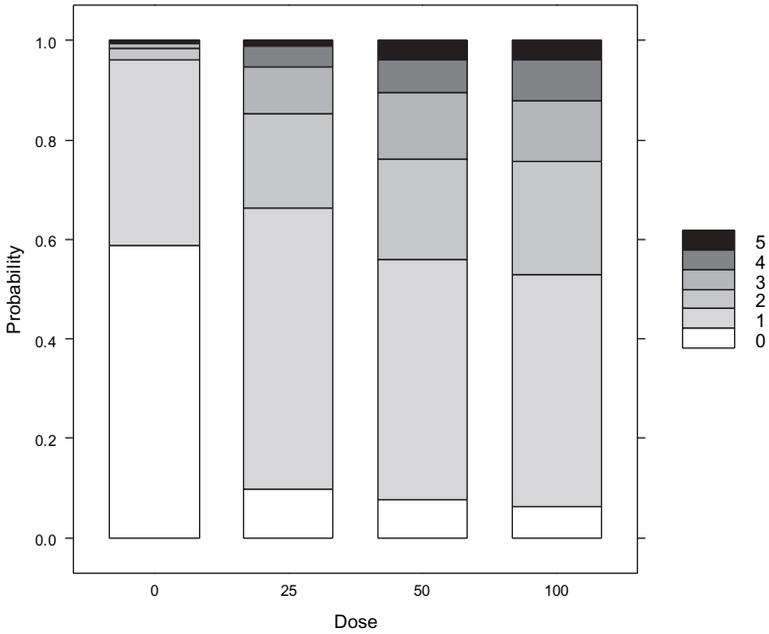
FIGURE 7.21   A stacked bar chart of the observed probabilities of the data versus the dose group. Each bar is divided into tiers that correspond to the probability of the observation category, ordered from top to bottom.

FIGURE 7.22   A stacked bar chart of the observed probabilities of the data versus the dose group. Each bar is divided into tiers that correspond to the probability of the observation category, ordered from top to bottom. The probability of each category is connected with a dashed line.
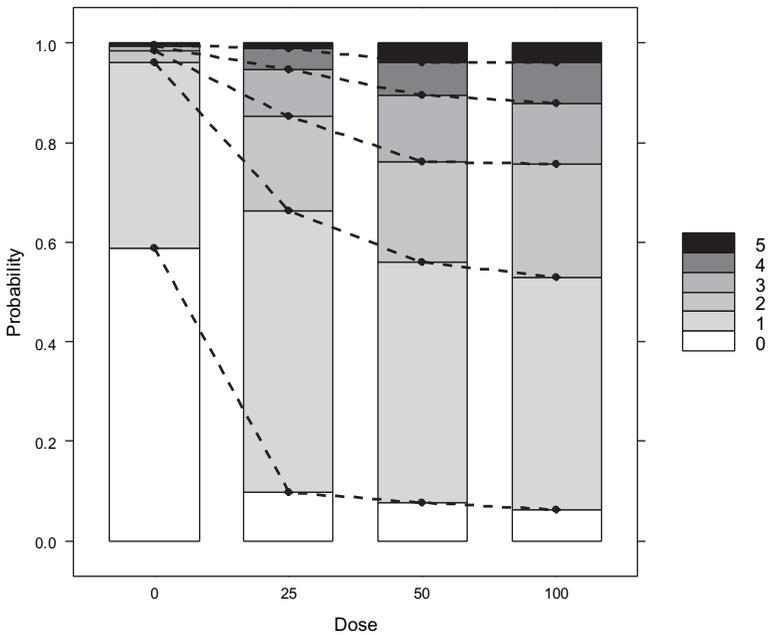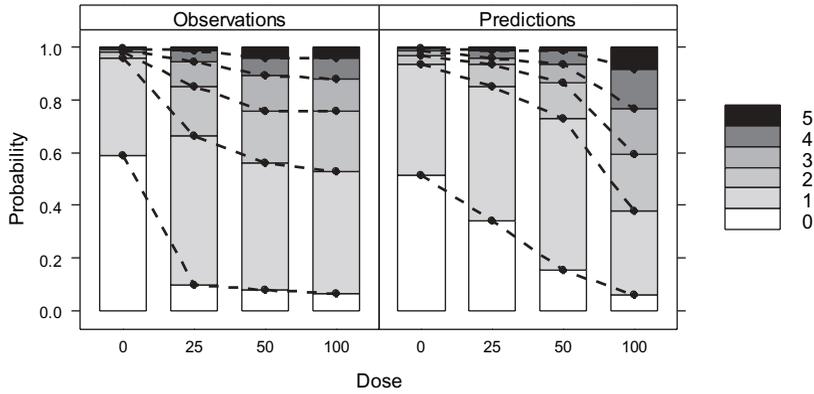
FIGURE 7.23   Stacked bar charts of the observed and predicted probabilities resulting from a fit of a model in which dose was included linearly in the logit (the "wrong" model).
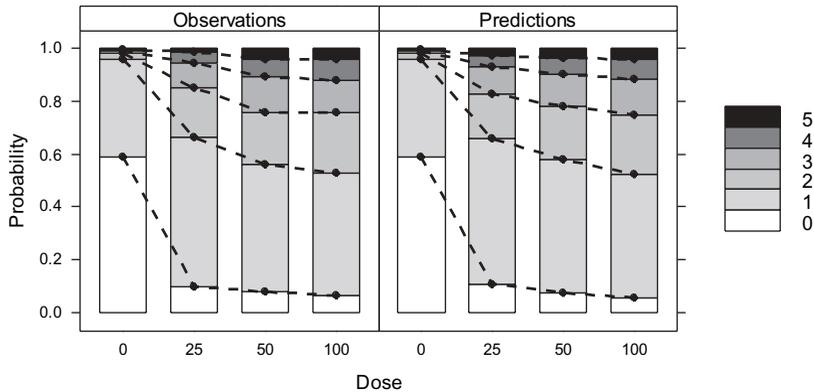

FIGURE 7.24   Stacked bar charts of the observed and predicted probabilities resulting from a fit of a model in which dose was included according to an E_{max} model in the logit (the "right" model). The predictions are based on 25 simulated data sets.

over all simulated data. In Figure 7.24 this has been done for the correct model, in which dose was added according to an E_{max} model. The predictions are based on 25 simulated data sets, of which the simulated values are pooled in the computation of the probabilities. In this example the extra simulated data sets did not matter much (not shown) but it may for smaller and/or more variable data sets.

Another fact to keep in mind is that the observed data is also a single random realization of a probability distribution. This means that the model may provide a good description of this distribution but that there is an apparent misfit only because of sample variability. Again, the solution is to simulate multiple data sets from the model and use the multiply predicted probabilities to construct prediction intervals. If the model provides an appropriate description of the observed data, we would expect the lines connecting the cumulative probabilities to be included in
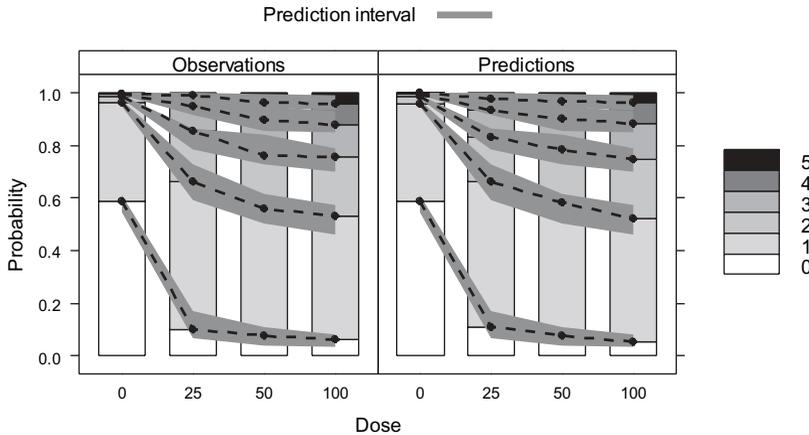
FIGURE 7.25   The same graph as in Figure 7.24 except that a prediction interval, based on the 25 simulated data sets, has been superposed in both panels.

the prediction intervals. This is exemplified in Figure 7.25. The graph is the same as in Figure 7.24 except that a prediction interval based on the 25 simulated data sets has been superposed in both panels (the shaded areas). The prediction intervals are computed as the appropriate percentiles from the 25 simulated values for each point in the lines based on the observed data. In both panels, the curves are centered in the shaded areas. This is expected, of course, for the predictions. With the wrong model the curves connecting the cumulative probabilities were way outside the shaded areas (not shown).

If we accept that the corresponding tiers in each of the panels in, for example, Figure 7.25 constitute an observation–prediction pair, it is also possible to investigate goodness of fit with more familiar plot types, for example, plotting the observed versus the predicted probability. (This approach is less useful if the number of observation categories is small.) This is shown in Figures 7.26 and 7.27 for the misspecified and correct models, respectively, and clearly the correct model provides a much closer fit. However, if we still are in doubt about the model appropriateness we can resort to simulations. The approach is identical to the one described above. Use the proposed model to simulate data, fit the model to the data, and produce predictions and plots in the same way that was used for the real data. Another approach is to use the simulations to construct prediction intervals for the points in Figures 7.26 and 7.27 and visualize them as horizontal error bars (not shown).

7.5   AFTER ANALYSIS

After the final model has been defined, tested, and checked according to all relevant means, the pharmacometrician needs to communicate the results to colleagues, peers, authorities, and so on. Although parameter estimates can be interesting, graphical displays play a major role in this phase.
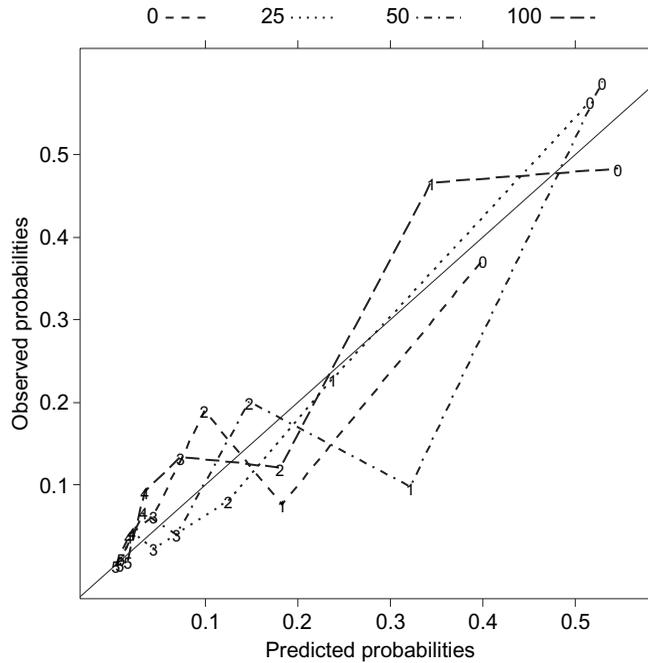
FIGURE 7.26 Shown are the observed probabilities for each category, given dose, plotted versus the corresponding predicted probabilities based on the wrong model. The solid diagonal line is the line of identity. The data points for each dose level are connected and each data point is labeled with the category number.

There are basically three distinct challenges to face. The first is to convince other pharmacometricians that the model adequately characterizes the data. This is usually done by an array of goodness of fit graphs, which can be quite technical if necessary since the intended audience is other modeling experts.

The second task involves defining the boundaries within which the model was derived and can be expected, without further motivation, to be an adequate description of the data. Within these bounds the model can replace the raw data (after all, the model is supposed to capture all salient features of the data), which will be useful when presenting the knowledge summarized by the developed model to nonpharmacometricians. The definition of these bounds can be based on the inclusion/exclusion criteria of the study or the realized covariate distribution. In the latter case, some of the exploratory plots from the "before-analysis phase" are useful.

The third task, and the one that is much too often overlooked, is the communication of the learnings of the modeling to subject matter experts, for example, project team members. This involves translating the model into quantities and pictures that nonpharmacometricians can relate to and that directly address the aim of the analysis. How this should be done has to be decided on a case by case basis depending on what the question of interest is. Despite this, we give three examples that we believe are of a more general nature.
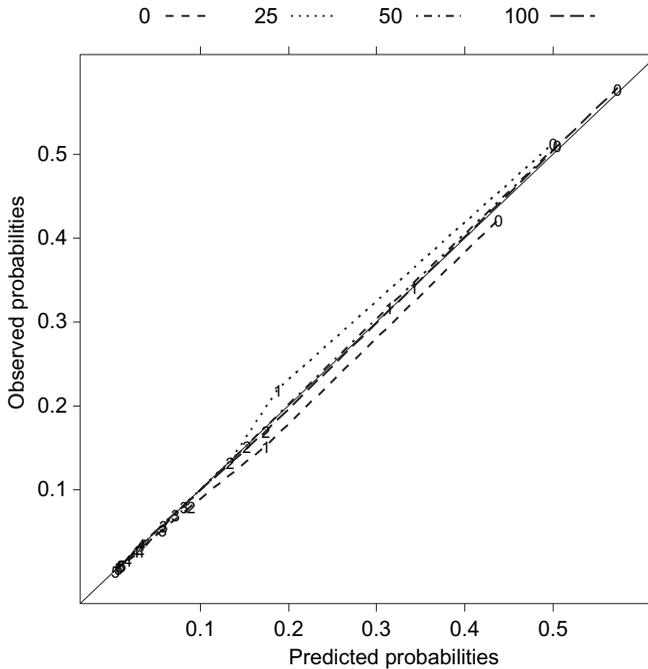
$$0 \; \text{-\;-\;-} \qquad 25 \; \cdots\cdots \qquad 50 \; \text{-\;\cdot\;-\;\cdot} \qquad 100 \; \text{—\;-\;}$$

FIGURE 7.27 Shown are the observed probabilities for each category, given dose, plotted versus the corresponding predicted probabilities based on the correct model. The solid diagonal line is the line of identity. The data points for each dose level are connected and each data point is labeled with the category number.

7.5.1 Visualizing the Relative Contribution of Covariates in Explaining Variability

A common issue is to communicate the relative importance of covariates in the explanation of unexplained variability in parameters, exposure, or pharmacodynamic endpoints. Consider the following schematic piece of code:

```
sexcl    = 0
if SEX   = 0 then sexcl=θ₃
TVCL     = θ₁(1+θ₂(CRCL-90))(1+sexcl)
CL       = TVCL*eⁿ
```

θ_1 is the typical value of clearance for a male (SEX=0) individual with creatinine clearance (CRCL) of 90, θ_2 is the fractional change in CL for each unit of CRCL different from 90, and θ_3 is the fractional difference in CL for females. The prediction of the individual value of clearance (CL) is given by the last line, where η_1 accounts for the remaining variability after taking the covariates into account. η_2 is a symmetrically distributed, zero mean variable with a variance of ω^2.

The values of the parameters (θ_1–θ_3 and ω^2) are of course estimated by the modeling program. When presenting the results to nonpharmacometricians, for example,

the clinical development team, the relevant question is how much of the overall variability in clearance (or exposure or pharmacodynamic endpoint) is explained by the covariates and *is not* the actual values of the parameters. The size of the parameter estimates will of course have a part in the assessment of the relative contribution of the covariates to explain the overall variability, but it is also necessary to take the variability of the covariates and the unexplained variability into account. In other words, it is not enough to report only the parameter values.

An approximate relative explanatory contribution of the covariates to a particular parameter can be computed directly from the parameter estimates and the distribution of the covariates. Using the code above, a measure of the overall variability is given by `var(`θ_2`(CRCL-90))`^{2}`+var(1+sexcl)`^{2}`+` ω^2—in other words, the sum of the variances of the individual components in the expression for the parameter in question. Note that `CRCL` and `sexcl` are here used in a vector sense in contrast to the code above, where it is used in a scalar sense. The approximate relative contributions of each variability component can now easily be computed as the ratio of the individual component to the overall variability. This is exemplified in Figure 7.28.

In many instances the variability in clearance reflects the variability in exposure, which makes Figure 7.28 relevant. If the exposure measure is not mirroring clearance, for example, time above a certain concentration, or if the focus is on the variability in a pharmacodynamic endpoint, the measures of variability can be obtained through simulations from the final model(s).

7.5.2 Visualizing the Effect of Individualized Dosing

Once the important covariates have been identified, it is necessary to decide if they can be used for individualized dosing and if this will lead to meaningful reduction in the variability in the exposure or effect. An informative display of the effect of individualized dosing is shown in Figure 7.29. It shows the average concentration as the
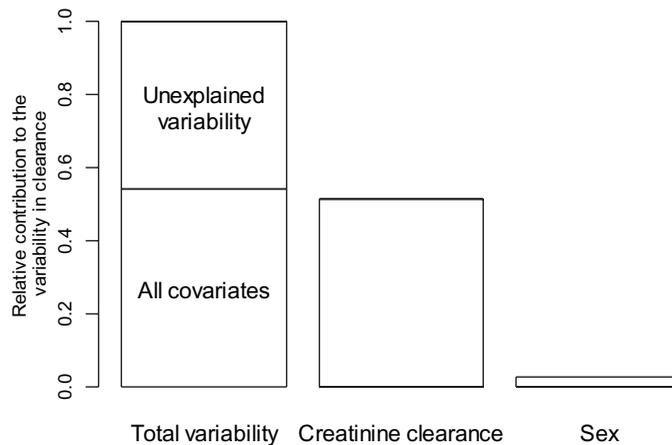

FIGURE 7.28   A bar plot visualizing the relative contributions of the covariates and unexplained variability to the total variability in clearance.
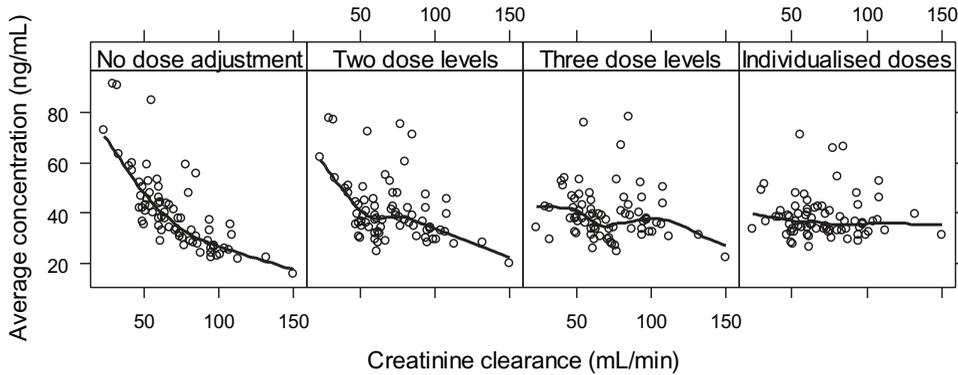
FIGURE 7.29 The graphs show the effect of dose individualization based on creatinine clearance. In the three left panels, one, two, and three dose levels were used, respectively. In the right panel, all individuals received their own dose based on creatinine clearance. See the text for details.

measure of exposure plotted versus the important covariate creatinine clearance. The data points are based on the model, the estimated η variables and the observed values of the covariate. If more than one covariate had been judged important, we could have used a multipanel conditioning version of the same type of display.

In Figure 7.29 the target concentration was arbitrarily set to the median observed concentration (the target needs to be determined on a drug by drug basis of course). The left panel shows the average concentration versus creatinine clearance if all individuals receive the same dose. The right panel, on the other hand, represents the extreme when all individuals receive an individualized dose given creatinine clearance. The two middle panels show what happens if each individual receives one out of two and three dose levels, respectively. The aim is to make the overall response as similar to the right panel as possible. In this case the three dose levels, with cutoff creatinine values at 35 and 80 mL/min, respectively, come quite close to the right panel while the two dose levels, with a cutoff at the median creatinine clearance (67 mL/min) do not. With more than one covariate to take into account, a trial and error approach to the question of appropriate cutoffs and dose increments may be quite inefficient. It is possible, however, to use the model to *estimate* the optimal cutoffs and dose increments, as described by Jönsson and Karlsson (18). Another possibility is to use tree-based models, at least if only the cutoffs are of interest (19).

7.5.3 Prediction and Confidence Intervals

A final task is to communicate variability and uncertainty. Given the multilevel nature of nonlinear mixed effects models, it is worth emphasizing the difference between these two. Variability is caused by true biological variability, making individuals different in various ways, while uncertainty is a measure of the (un)certainty in the estimated model parameters. One illustrative way of displaying this is shown in Figure 7.30. The graph shows the 95% prediction and confidence intervals around
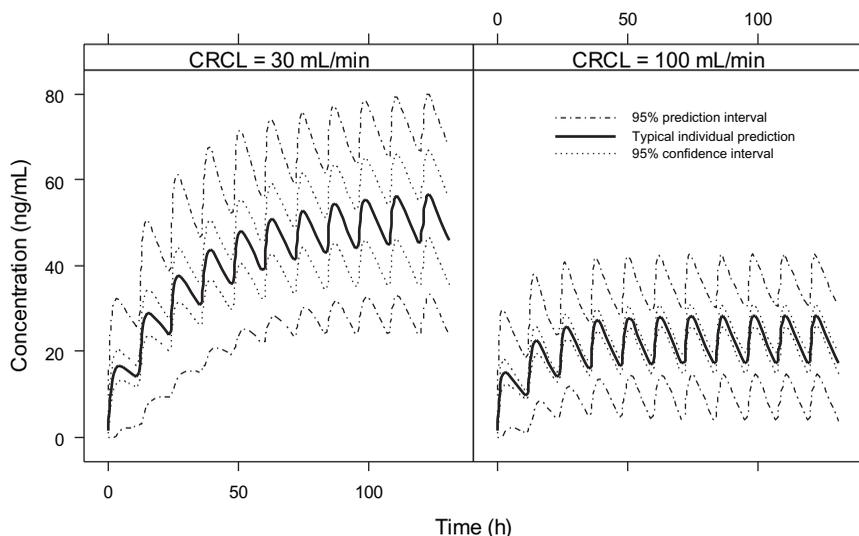
FIGURE 7.30   Shown is the predicted concentration time curves (solid lines) over 11 dosing intervals together with the 95% confidence (dotted lines) and prediction (dashed lines) intervals for a typical individual with poor renal function (left) and normal renal function (right).

the typical individual curve for individuals with a low renal function (30 mL/min) and normal renal function (100 mL/min) in the left and right panels, respectively. The 95% prediction interval is the interval in which we expect 95% of the observations from future individuals to fall and the 95% confidence interval around the concentration time profiles indicates the certainty in the predicted typical individual curve; that is, there is 95% certainty that the true concentration–time profile is included in this interval. How these intervals are computed is shown in Appendix 7.3.

By comparing the prediction and confidence intervals, we can make a judgment about the certainty in the predictions made from the model. For example, according to the right panel in Figure 7.30 we are relatively confident about the typical individual predictions compared to the prediction interval range. In the left panel we can see that the uncertainty in the predicted curve is greater, indicating that we can potentially gain certainty by collecting more data in this patient group.

7.6 SUMMARY

Graphs are useful in all phases of population PK/PD modeling. Before the analysis, the importance lies in data set checkout as well as exploratory analysis. During the analysis, graphical analysis is the mainstay in model diagnostics and guides model development. After the analysis, when the results need to be communicated, graphics can be used to transparently convey quite involved information.

We demonstrate that the hierarchical, variable, multidimensional, and potentially plentiful and categorical nature of population data can be handled in graphically

informative fashions. This is not to say that it is always easy, but it illustrates that it is a worthwhile exercise to become an expert user in whatever plotting program is at one's disposal and then exploit that expertise methodically.

REFERENCES

1. W. S. Cleveland. *Visualizing Data*. Hobart Press, Summit, NJ, 1993.

2. L. Brynne, M. O. Karlsson, and L. K. Paalzow, Concentration–effect relationship of l-propranolol and metoprolol in spontaneous hypertensive rats after exercise-induced tachycardia. *J Pharmacol Exp Ther* **286**:1152–1158 (1998).

3. L. Brynne, L. K. Paalzow, and M. O. Karlsson, Consequence of exercise on the cardio-vascular effects of l-propranolol in spontaneously hypertensive rats. *J Pharmacol Exp Ther* **294**:1201–1208 (2000).

4. M. O. Karlsson and L. B. Sheiner, The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Biopharm* **21**:735–750 (1993).

5. E. I. Ette, Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial. *Ann Pharmacother* **32**:818–828 (1998).

6. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

7. E. I. Ette, P. Williams, E. Fadiran, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).

8. E. N. Jonsson and M. O. Karlsson, Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* **58**:51–64 (1999).

9. M. O. Karlsson, E. N. Jonsson, C. G. Wiltse, and J. R. Wade, Assumption testing in population pharmacokinetic models: illustrated with an analysis of moxonidine data from congestive heart failure patients. *J Pharmacokinet Biopharm* **26**:207–246 (1998).

10. U. Wahlby, A. H. Thomson, P. A. Milligan, and M. O. Karlsson, Models for time-varying covariates in population pharmacokinetic–pharmacodynamic analysis. *Br J Clin Pharmacol* **58**:367–377 (2004).

11. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. Globomax, Inc., Baltimore, MD, 1989–1998.

12. M. O. Karlsson, S. L. Beal, and L. B. Sheiner, Three new residual error models for population PK/PD analyses. *J Pharmacokinet Biopharm* **23**:651–672 (1995).

13. R. J. Carroll and D. Ruppert, Transformation and weighting in regression. Chapman & Hall, New York, 1988.

14. J. Gabrielsson and D. Weiner, Pharmacokinetic and pharmacodynamic data analysis. Swedish Pharmaceutical Press, Stockholm, Sweden, 1997.

15. E. H. Cox, C. Veyrat-Follet, S. L. Beal, E. Fuseau, S. Kenkare, and L. B. Sheiner, A population pharmacokinetic–pharmacodynamic analysis of repeated measures time-to-event pharmacodynamic responses: the antiemetic effect of ondansetron. *J Pharmacokinet Biopharm* **27**:625–644 (1999).

16. L. B. Sheiner, A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther* **56**:309–322 (1994).

17. P. H. Zingmark, M. Ekblom, T. Odergren, T. Ashwood, P. Lyden, M. O. Karlsson, and E. N. Jonsson, Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol* **56**:173–183 (2003).

18. S. Jönsson and M. O. Karlsson, A rational approach for selection of optimal covariate-based dosing strategies. *Clin Pharmacol Ther* **73**:7–19 (2003).

19. M. Chambers and J. Hastie, *Statistical Models in S*. Chapman & Hall, New York, 1993.

APPENDIX 7.1   OBTAINING INDIVIDUAL PREDICTIONS AND RESIDUALS

Individual predictions (IPRED) and residuals (IWRES) are not defined by NONMEM but may easily be obtained through the following fragment code in the NM-TRAN control stream.

```
$ERROR
  IPRED           = F
  IRES            = DV-IPRED
  W               = THETA(y)*IPRED
  DEL             = 0
  IF(W.EQ.0) DEL  = 1
  IWRES           = IRES/(W+DEL)
  Y               = F+EPS(1)*W

$SIGMA 1 FIX
$EST POSTHOC
```

The above code specifies a constant CV residual error model and THETA(y) will be the standard deviation of the residual error (note the $SIGMA 1 FIX). The DEL variable protects from division by zero in the line with IWRES. IPRED and IWRES can be output in a table file.

To implement a slope-intercept residual error model is as simple as replacing the W= line above with

```
W=SQRT(THETA(x)**2+THETA(y)**2*IPRED**2)
```

where THETA(x) will be the standard deviation of the additive residual error component and THETA(y) will be the standard deviation of the proportional component. An additive residual error model is implemented by changing the W= line to W=THETA(x), where THETA(x) is the standard deviation of the residual error.

APPENDIX 7.2   OBTAINING STANDARD ERRORS FOR η

Standard errors (SE) for individual η values are not provided in the standard NONMEM output. They can be obtained, however, by using the technique described below.

The first step is to create a new data file with one more column and a number of extra data records per individual. The extra records are the same as any observation record except that zeros are inserted in the DV column. The number of extra records per individual should be the same as the number of η values in the model. The new column signals the type of record and should have one value for the original

observations and one unique value for each new record. This is exemplified below
for the first two individuals in the PK data set.

ID	TIME	DV	AMT	SS	II	CRCL	TYPE
1	0.0	0	500	0	0	67.1	0
1	12.0	0	500	0	0	67.1	0
1	24.0	0	500	0	0	67.1	0
1	36.0	0	500	0	0	67.1	0
1	47.0	0	500	0	0	58.1	0
1	49.5	4.96	0	0	0	58.1	1
1	50.8	3.46	0	0	0	58.1	1
1	51.3	4.14	0	0	0	58.1	1
1	52.3	3.89	0	0	0	58.1	1
1	54.3	3.50	0	0	0	58.1	1
1	56.6	3.32	0	0	0	58.1	1
1	58.3	2.86	0	0	0	58.1	1
1	62.5	3.15	0	0	0	58.1	1
1	**63.0**	**0**	**0**	**0**	**0**	**58.1**	**2**
1	**63.0**	**0**	**0**	**0**	**0**	**58.1**	**3**
1	**63.0**	**0**	**0**	**0**	**0**	**58.1**	**4**
2	0.0	0	500	0	0	82.2	0
2	2.2	2.62	0	0	0	82.2	1
2	**2.2**	**0**	**0**	**0**	**0**	**82.2**	**2**
2	**2.2**	**0**	**0**	**0**	**0**	**82.2**	**3**
2	**2.2**	**0**	**0**	**0**	**0**	**82.2**	**4**

The boldfaced records have been added to the original data set, as has the TYPE
column.

   The model file also needs to be changed and an example is given below. All
changes from the original model file are boldfaced.

   The most obvious change is that one $PROBLEM statement for each individual
in the data set (in this case two) has to be created. Each $DATA statement in each
$PROBLEM has to contain an NREC option specifying the number of records for the
corresponding individual. In $PK, which can only be present in the first $PROBLEM,
the ETA() variables should be replaced by THETA() variables. The estimates of these
THETAs in the NONMEM output file will be the individual η values. In $ERROR,
which can also only be present in the first $PROBLEM, IF statements should be
added, one for each value in the new column in the data set (TYPE). With TYPE=1
(in this case), Y is set to the model prediction. Note that the EPS() variables have
been replaced by ETA() variables, whose variances are fixed to the corresponding
σ^2 estimates from the population model. With TYPE=2, 3, or 4, the Y is set to the
THETA() that replaces the ETA() in $PK, plus ETA() variables whose variances are
fixed to the corresponding ω^2 estimates from the population model. The $COVARI-
ANCE record needs to specify the option MATRIX=R, as is appropriate for the estima-
tion of SEs in single individual fits. Finally, one $PROBLEM needs to be added for
each remaining individual in the data set. Note that $SUBROUTINE, $PK, and $ERROR
are not allowed in $PROBLEMs following the first. For more detailed descriptions

of the various aspects in this control stream, please refer to the *NONMEM Users Guides* (11).

```
$PROB Estimating individual etas to obtain SEs: ID=1
$INPUT ID TIME DV AMT SS II CRCL TYPE
$DATA indpardata.prn NREC=16 IGNORE=@
$SUBROUTINE ADVAN2 TRANS2
$PK

  TVCL = THETA(1)*(1+THETA(4)*(CRCL-90))
  TVV  = THETA(2)
  TVKA = THETA(3)

  CL   = TVCL*EXP(THETA(5)) ; ETA replaced by THETA
  V    = TVV *EXP(THETA(6)) ;            -"-
  KA   = TVKA*EXP(THETA(7)) ;            -"-

  S2=V

$ERROR

  IPRED = F
  IF(TYPE.LE.1) THEN ; If regular observation
    Y = IPRED+IPRED*ETA(1)+ETA(2)
  ENDIF

; One if-statement for each new record in the data file
  IF(TYPE.EQ.2) Y = THETA(5)+ETA(3)
  IF(TYPE.EQ.3) Y = THETA(6)+ETA(4)
  IF(TYPE.EQ.4) Y = THETA(7)+ETA(5)

; Parameter estimates from the population model
$THETA 16.5    FIX ; THETA(1)
$THETA 263     FIX ; THETA(2)
$THETA 0.825   FIX ; THETA(3)
$THETA 0.00945 FIX ; THETA(4)

; Parameters to be estimated=the ETAs
$THETA 0.01          ;5 ETA on CL
$THETA 0.01          ;6 ETA on V
$THETA 0.01          ;7 ETA on KA

; Variance parameter estimates from the population model
$OMEGA 0.027  FIX  ;1 Proportional residual error
$OMEGA 0.132  FIX  ;2 Additive residual error
$OMEGA 0.0471 FIX  ;3 Omega^2 CL
$OMEGA 0.189  FIX  ;4 Omega^2 V
$OMEGA 0.572  FIX  ;5 Omega^2 KA
```

```
$EST
; Need to use MATRIX=R for single individual data
$COV MATRIX=R


$PROB Estimating individual etas to obtain SEs: ID=2
$INPUT ID TIME DV AMT SS II CRCL TYPE
$DATA indpardata.prn NREC=5 IGNORE=@

$THETA 16.5    FIX ; THETA(1)
$THETA 263     FIX ; THETA(2)
$THETA 0.825   FIX ; THETA(3)
$THETA 0.00945 FIX ; THETA(4)

$THETA 0.01         ;5 ETA on CL
$THETA 0.01         ;6 ETA on V
$THETA 0.01         ;7 ETA on KA


$OMEGA 0.027  FIX  ;1 Proportional residual error
$OMEGA 0.132  FIX  ;2 Additive residual error
$OMEGA 0.0471 FIX  ;3 Omega^2 CL
$OMEGA 0.189  FIX  ;4 Omega^2 V
$OMEGA 0.572  FIX  ;5 Omega^2 KA

$EST
$COV MATRIX=R
```

APPENDIX 7.3   CONFIDENCE AND PREDICTION INTERVALS

The computation of the prediction and confidence intervals around a predicted concentration time profile is based on the so-called delta method.[1] The method is based on a first-order like approximation.

The idea is to compute the variance of the typical individual predictions based on variability of the η values:

$$\mathrm{Var}\left(\hat{\bar{y}}\right) = \sum_{n=1}^{p}\sum_{m=1}^{p} \frac{\delta\hat{\bar{y}}}{\delta\eta_n} \frac{\delta\hat{\bar{y}}}{\delta\eta_m} \mathrm{Cov}(\eta_n, \eta_m) \tag{7.8}$$

where p is the number of η values in the model. The individual- and time-specific subscripts (i and j, respectively) have been suppressed for notational ease.

To compute 95% prediction intervals, Eq. (7.8) is implemented in NONMEM in the following way:

```
$PROB Code for computing approximate prediction intervals
$INPUT ID TIME DV AMT CRCL
$DATA madeup.dta IGNORE=@
```

---

[1]C. R. Rao, *Linear Statistical Inference and Its Application*. Wiley, Hoboken, NJ, 1973.

```
$SUBROUTINE ADVAN2 TRANS2
$ABBREV COMRES=2
$PK
  TVCL = THETA(1)*(1+THETA(4)*(CRCL-90))
  TVV  = THETA(2)
  TVKA = THETA(3)

  CL   = TVCL*EXP(ETA(1))
  V    = TVV *EXP(ETA(2))
  KA   = TVKA*EXP(ETA(3))
  S2=V
$ERROR
"FIRST
" COMMON /rocm6/THETAF(40),OMEGAF(30,30),SIGMAF(30,30)

  IPRED = F
  IRES  = DV-IPRED
  Y     = IPRED+IPRED*EPS(1)+EPS(2)

" LAST
" IF(COMACT.EQ.1) THEN
" VARCP = G(1,1)**2*OMEGAF(1,1)
" VARCP = G(2,1)**2*OMEGAF(2,2)+VARCP
" VARCP = G(3,1)**2*OMEGAF(3,3)+VARCP
" VARCP = G(1,1)*G(2,1)*OMEGAF(1,2)+VARCP
" VARCP = G(1,1)*G(3,1)*OMEGAF(1,3)+VARCP
" VARCP = G(2,1)*G(3,1)*OMEGAF(2,3)+VARCP
" VARCP = SIGMAF(1,1) + SIGMAF(2,2)+ VARCP
" SDCP  = SQRT(VARCP)
" COM(1)= IPRED+1.96*SDCP
" COM(2)= IPRED-1.96*SDCP
" ENDIF

$THETA (0,16.5)     ;CL Set to the final estimate
$THETA (0,263)      ;V               -"-
$THETA (0,0.825)    ;KA              -"-
$THETA (0,0.00945)  ;CRCL            -"-
$OMEGA 0.0474       ;CL              -"-
$OMEGA 0.190        ;V               -"-
$OMEGA 0.572        ;KA              -"-
$SIGMA 0.027        ;SIGMA(1)        -"-
$SIGMA 0.132        ;SIGMA(2)        -"-
$EST MAXEVAL=0
$TABLE ID TIME IPRED COM(1)=UPP COM(2)=DOWN
       NOPRINT ONEHEADER FILE=pinttab
```

The variable `G()` available in NONMEM verbatim code (lines starting with ")
holds the partial derivatives of the predictions from `$PK` with respect to the η values.

OMEGAF and SIGMAF contain the estimates of OMEGA and SIGMA, which in this case will be the same as the values specified on the $OMEGA and $SIGMA rows since MAXEVAL is set to 0. The limits of the prediction intervals are stored in the variables COM(1) and COM(2), which are tabulated. The input data file has to include all time points for which the interval is to be computed (including any doses).

   Computing the 95% confidence intervals is similar to the computation of the 95% prediction intervals *but not identical*.

```
$PROB Code for creating approximate confidence intervals
$INPUT ID TIME DV AMT CRCL
$DATA madeup.dta IGNORE=@
$SUBROUTINE ADVAN2 TRANS2
$ABBREV COMRES=2
$PK
; The ETAs should be entered additively
  TVCL = THETA(1)*(1+(THETA(4)+ETA(4))*(CRCL-90))
  TVV  = THETA(2)
  TVKA = THETA(3)

  CL   = TVCL+ETA(1)
  V    = TVV +ETA(2)
  KA   = TVKA+ETA(3)
  S2   = V

$ERROR
"FIRST
" COMMON /rocm6/THETAF(40),OMEGAF(30,30),SIGMAF(30,30)

  IPRED = F
  IRES  = DV-IPRED
  Y     = IPRED+IPRED*EPS(1)+EPS(2)

" LAST
" IF(COMACT.EQ.1) THEN
"    VARCP = G(1,1)**2*OMEGAF(1,1)
"    VARCP = G(2,1)**2*OMEGAF(2,2)+VARCP
"    VARCP = G(3,1)**2*OMEGAF(3,3)+VARCP
"    VARCP = G(4,1)**2*OMEGAF(4,4)+VARCP
"    VARCP = G(1,1)*G(2,1)*OMEGAF(1,2)+VARCP
"    VARCP = G(1,1)*G(3,1)*OMEGAF(1,3)+VARCP
"    VARCP = G(1,1)*G(4,1)*OMEGAF(1,4)+VARCP
"    VARCP = G(2,1)*G(3,1)*OMEGAF(2,3)+VARCP
"    VARCP = G(2,1)*G(4,1)*OMEGAF(2,4)+VARCP
"    VARCP = G(3,1)*G(4,1)*OMEGAF(3,4)+VARCP
"    SDCP  = SQRT(VARCP)
"    COM(1) = IPRED+1.96*SDCP
"    COM(2) = IPRED-1.96*SDCP
" ENDIF
```

```
$THETA (0,16.5)     ;CL Set to the final estimate
$THETA (0,263)      ;V              -"-
$THETA (0,0.825)    ;KA             -"-
$THETA (0,0.00945)  ;CRCL           -"-


; The values in the OMEGA BLOCK are taken from the variance-
; covariance matrix of the estimate in the NONMEM output file
$OMEGA BLOCK(4) 0.541
                -3.81 940
                -0.0231 0.713 0.0213
                .000504 -.00358 -.000015 .0000008


$SIGMA 0.027 0.132 ; These values are irrelevant
$EST MAXEVAL=0
$TABLE ID TIME IPRED COM(1)=UPP COM(2)=DOWN
  NOPRINT ONEHEADER FILE=citab
```

The idea here is to use the values for the THETAs in the variance–covariance matrix of the estimate in the $OMEGA BLOCK, which are available in the NONMEM output file after a successful $COVARIANCE step. It is necessary to use additive models for the η values as well as adding an η on the parameter for the creatinine clearance relation THETA(4). Note that the value in $SIGMA is not used in the computations and can be set to anything. For further details of the code, please consult the NONMEM manuals and nmhelp (the online help system distributed with NONMEM). More precise reflections of the confidence and prediction intervals can be obtained by multiple simulations from the final model and suitable dosing/observation patterns followed by creation of piecewise prediction/confidence intervals from the simulated observations.

The Epistemology of Pharmacometrics

PAUL J. WILLIAMS, YONG HO KIM, and ENE I. ETTE

8.1 INTRODUCTION

Epistemology is the field of study that addresses the theory of knowledge, that we engage in to determine belief, truth, and knowledge. Ultimately one engages in the pursuit of knowledge and belief to make wise decisions, and making wise decisions in the realm of pharmacometrics is of extreme importance because of its applicability to optimize pharmacotherapy and drug development. The kinds of wise decisions one wishes to make most often involve dosing strategies or patient selection whether ultimately for labeling or for pivotal studies leading to regulatory approval of a therapeutic agent. Thus, having the correct beliefs about the pharmacometric (PM) properties of a drug is not simply a matter of intellectual importance but of practical importance. Knowledge is a properly justified belief. It can be dangerous to claim to have knowledge that is not justified and one must recognize that there are differing levels of risk for false belief depending on the consequences.

When generating PM knowledge, one is not attempting to state the exact truth but one wishes to create knowledge for which there is some degree of correspondence with the truth. The degree of correspondence that one seeks between the PM knowledge and the truth is determined by the intended use of the PM knowledge and the consequences of false belief. PM models are used to summarize what is known about the biology (or state) of the system being investigated. Thus, one must ask if the model is appropriate for its intended use. Therefore, the pivotal event for establishing model appropriateness is stating clearly how the model will be applied and what the consequences of false belief are. Model appropriateness has been discussed in detail previously (1).

PM models have been used to answer questions such as: What is an optimal dosing strategy? Should dosing strategies be based on size, renal function, liver function, genomics, and so on? What should the first-time dose in humans be? How should a clinical confirming study be designed? They have been applied to preclinical data for estimating animal exposure and response (2); for allometric scaling of pharmacokinetics from animals to humans (2); for determining first-time-in-human

(FTIH) dosing strategies by linking pharmacokinetics and pharmacodynamics (2); for understanding covariate level impact on human pharmacokinetics and dosing (3, 4); and as an exercise executed for drug approval.

The actual model estimation and development should proceed, using sound methods (4). The nature of the manner in which a PM model will be applied to solve the problem must be accounted for during the entire modeling process. The intended use of a model ought to influence the attitude and modeling approaches used by the pharmacometrician at the various stages of the modeling process. The intended use would determine what covariates are considered important, which parameters are of primary concern, and what the extent and method of model evaluation and validation should be. Thus, how the model will be used to solve the problem drives the modeling process from model development, through model evaluation, to model validation that establishes whether a model is appropriate.

8.2 DEFINITIONS

The absence of the terms "model qualification," "model verification," "model accreditation," and "credible model" as applied to PM models deserves explanation. The definition for qualified is "having complied with specific requirements or precedent conditions." Model qualification would imply that some specific objective standard must be met for a model to be "qualified," that the standards are the same for all models, and that there would be no alternative approaches to "qualifying" the model. Within the realm of pharmacometric (PM) models, these specific precedent conditions have not been stated, and it would be impossible and unreasonable to have a set of specific objective conditions that would apply to all models.

Accreditation is a term that should not be used for PM models. The same criticisms apply to accreditation as to qualification. In addition to the criticisms applied to model qualification, documentation is a further necessary step for model accreditation. Documentation includes verification, which includes assessing the conceptual models and verifying the computer model. The substitution of model accreditation for appropriateness across disciplines could lead to cross-discipline confusion. Therefore, model accreditation should not serve as a substitute for model appropriateness (5).

The term "model verification" has two problems. First, "model verification" has been applied in other disciplines to mean that the computer program used to execute the model and its implementation are correct. Therefore, if applied to PM modeling, model verification would cause some cross-discipline confusion. Also, "verification" comes from the Latin *veritas*, which means "truth"; to verify means to establish the truth, accuracy, or reality of something. Models are approximations to the truth or to the real systems that they describe and are not the truth, as would be implied by the term "verification" (5, 6). Therefore, verification is not synonymous with appropriateness.

Credible models are those created in the absence of data and based on expert opinion (6). For example, to obtain permission for nuclear waste disposal in the waste isolation pilot plant near Carlsbad, New Mexico, a model was developed at Sandia National Laboratories in the absence of data on nuclear waste facilities and was based on expert opinion. In the end, the Environmental Protection Agency will

accept the plan only if the model is credible and the model's output shows acceptable results. For PM models the only time a credible model is needed is before in vivo experiments are conducted. Therefore, we seldom function with credible models and the terminology credible will not be used here (6, 7).

Appropriateness means suitability or aptness; that is the concept we are attempting to convey here. Model appropriateness is presented in such a way as to be compatible with terminology used in the statistical and other modeling literature, which are foundational for PM modeling. In subsequent sections, we present the steps in establishing model appropriateness, approaches to model evaluation and validation, metrics used in model validation, and an application example to illustrate the principles involved in establishing model appropriateness.

8.3   MODEL APPROPRIATENESS

8.3.1   Steps for Establishment of Model Appropriateness

A flow diagram for establishing an appropriate model is presented in Figure 8.1. The steps to be executed to establish an appropriate model are noted in the figure. The method of model estimation, development, evaluation, and validation should be prescribed prior to model development. An overview of the necessary elements establishing an epistemologically appropriate model are presented here and in more detail elsewhere (1).
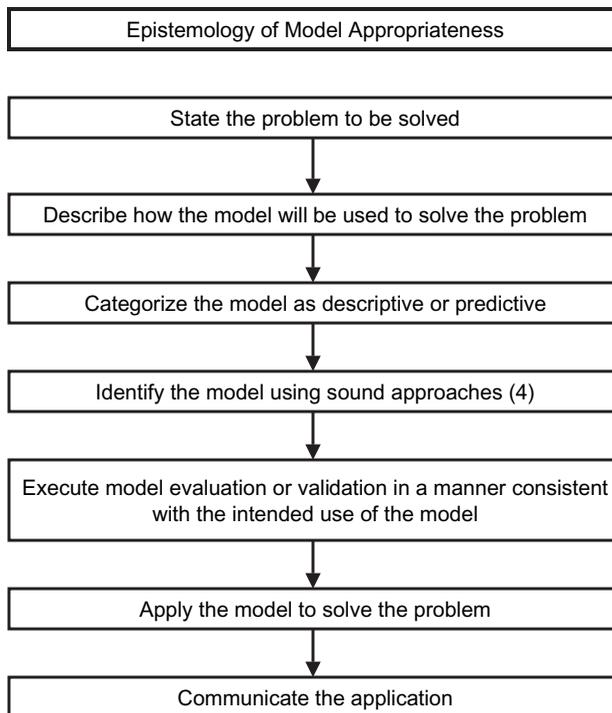

FIGURE 8.1   Overview of model appropriateness.

8.3.2 Types of Models

PM models are either descriptive or predictive. Descriptive models are those PM models applied to the patients' (data) from which the models have been derived or estimated and for which there is no intent to extrapolate or apply the model to a population other than the population from which the model was derived. That is, descriptive models explain the variability of the pharmacometrics of the drug and, in this case, are used as an empirical and numerical summary of information about PM variability in the population studied and would include all features of the population found to be important covariates. For example, a descriptive model may be used to explain a higher incidence of adverse effects of a drug in one subgroup of subjects versus another. Perhaps a drug is primarily eliminated by the kidney and that subgroup of subjects with impaired renal function has a lower clearance and therefore greater exposure to the drug than the subgroup without impaired renal function. Or, the group of subjects with impaired renal function had a greater incidence of adverse effects than the group with normal function. One possible explanation for the greater incidence of adverse effects in the impaired renal function group is that they had a greater exposure to the drug resulting in greater toxicity. This would be adequate as a key source of information that would help to explain subgroup differences in response and would be a descriptive model because it was applied only to the subjects from whom it was derived.

When a model is used for descriptive purposes, goodness-of-fit, reliability, and stability, the components of model evaluation must be assessed. Model evaluation should be done in a manner consistent with the intended application of the PM model. The reliability of the analysis results can be checked by carefully examining diagnostic plots, key parameter estimates, standard errors, case deletion diagnostics (7–9), and/or sensitivity analysis as may seem appropriate. Confidence intervals (standard errors) for parameters may be checked using nonparametric techniques, such as the jackknife and bootstrapping, or the profile likelihood method. Model stability to determine whether the covariates in the PM model are those that should be tested for inclusion in the model can be checked using the bootstrap (9).

When models are not checked for stability, spurious covariates may be included in the model because of leverage or influence data that may have their source in only a few subjects. Small changes in a data set may result in the selection of different covariates for a PM model when a model is not checked for stability.

A second class of PM models are used for predictive purposes. Predictive models are intended to have some impact or application to patients or subjects from whom data has not been obtained. Although these models contain descriptive components, they are used to answer "what if" questions about the effects of changes in the covariates of the model. For a predictive model, if one is interested in evaluating the effect of several dosing strategies on outcomes for a pivotal Phase 3 study via simulation, then the distribution of the PK parameters becomes very important. Another example would be the effect of subpopulation differences on a dosage strategy proposal.

When PM models are developed for predictive purposes, much stronger assumptions are made about the relationship between the underlying population, from whom the data were collected, and the predicted results. One is asking for correspondence of behavior outside the range over which one has actual empirical

evidence. In other words, one is intending to apply the model to patients or subjects from whom it was not estimated. In the descriptive PM model, what was observed is important; in the predictive PM model, the behavior of the model is important.

For predictive purposes, a PM model should be validated in consistence with the intended use of the model. Valid means to be well grounded, convincing, sound, or "having such force as to compel acceptance" (this is the sense of the use of this term in the statistical and modeling literature) but does not mean to be true (10, 11). When validating, it is always important to remember the dictum from Box (12): "All models are wrong, some are useful." Therefore, asking whether one's model is true or false is not appropriate; rather, one should ask whether the model's deficiencies have a noticeable effect on the substantive inferences. Model validation results in confidence that the model does not have deficiencies that will result in it not being applicable for its intended use.

8.4   MODEL IDENTIFICATION

Systematic (or fixed) effects in a clinical trial data set are blurred by other variations of a more haphazard nature. A haphazard variation is usually described in pharmacostatistical terms as random effects. PM models contain both fixed and random effects elements. A value of a PM model is that it provides a summary of the data in terms of the fixed effects and the nature and magnitude of the random or unexplained variation. Estimating PM models entails examining the data intelligently and this demands formulating models that are thought to be capable of not only characterizing the systematic variation in the data analyzed but also describing patterns in similar data collected elsewhere if the PM model developed is to be transportable (1). Thus, the development of a PM model begins with the identification of the model with which a population model is to be developed.

To identify the model appropriate for characterizing the data at hand, the pharmacometrician begins with ideas, experience, literature information (if it exists), knowledge of models developed from the drug from prior studies and in the case of a first-time-in-human study, knowledge gained from models developed from animal(s) to extrapolate to humans (especially when nonlinear mixed effects modeling was used), and of course the data to be analyzed. The process will vary from problem to problem, and there are no general rules. This not withstanding, we present herein a general approach to model identification. Although the thought process presented here will be illustrated with pharmacokinetics, the general principles are applicable to pharmacodynamics as well.

8.4.1   Identification of a Base Model

1. *Plot of Concentration–Time Profile for Population Data Set.* It is advisable to plot the concentration–time profile of data for the entire population studied to reveal patterns and structure in the data (4) (see Chapter 14). Sometimes it might be necessary to drill into the plot of the profiles to examine different aspects of the profile. For instance, characterizing absorption may require examining the absorption phase of the concentration–time profile plot closely. This would enable the answering of questions such as: Is there a lag time in absorption? Is

the absorption rapid, gradual, or slow? Are there different absorption patterns for different groups of subjects? Knowledge gained from such an examination by answering the above questions would inform the type of absorption model that could be used to model the absorption profile observed (see Chapter 13 for modeling absorption).

2. *Begin with the Simplest Basic Structural Model.* This should be the natural starting point after examining the nature of the profile. A basic structural model is a model (e.g., one-compartment PK model) without covariates. A complex initial model makes it difficult to see where changes should be made. Although one may aim to develop a model that reflects "reality," simplicity is a paramount requirement. Sometimes prior knowledge may inform a more complex structural model than a simple one—two-compartment versus a one-compartment PK model, for instance—but the data at hand may not support the more complex model. A less complex model may be implemented for a population analysis rather than the model established in earlier studies. This use of a less complex or "incomplete" model has been described in the literature in another context as the "minimal model" approach (13). This idea, although not explicitly stated, was present in a paper describing the population analysis of the pharmacokinetics of digoxin in patients (14). Although the pharmacokinetics of digoxin are known to be described by a multicompartment model, the authors chose to use a one-compartment model to describe the data. This is because all concentration measurements had been obtained in the "postdistribution" phase, as concentrations measured early after the doses were considered to be of little clinical relevance.

It is important to note that the simplest base model for preterm or full-term infants sometimes may include weight for the model to be stable before further modeling can proceed.

3. *Optimize the Structural Model.* It is important to ensure that the structural model describes the underlying patterns in the data. If the data speaks to the existence of two clearly distinct absorption profiles, then a mixture model should be tested to ensure that the absorption phase of the profile is well characterized. If absorption, for example, could be better characterized with sequential first-order absorption models instead of a simple first-order model, the appropriate model should be used to eliminate bias due to model misspecification.

4. *Include Intersubject Variability on Model Parameters and a Residual Error Model.* Modeling intersubject random effect with the basic structural model is essential for a PM model. Some prefer not to include η values (intersubject variability) on all parameters, but do so as more confidence is gained with the base model. Some include η values on all parameters and only drop them if the model estimation indicates that they are unnecessary. The choice is left to the pharmacometrician to go with the approach he/she is most comfortable with. A lognormal model should be a starting point for modeling intersubject variability. This is because biological variation by nature is lognormal.

In modeling intersubject variability, it is advisable to start with the diagonal elements of the covariance matrix instead of starting with the full covariance matrix. If NONMEM is used for modeling, the posthoc parameter (or η) values should be subjected to a pairs plot to determine if there are any relationships (correlations) between parameters. A correlation test should be performed for parameters that

appear to be correlated with each other. Moderate ($0.5 \leq r \leq 0.74$) to high ($r \geq 0.75$) correlations should be accounted for using the omega block code. The likelihood ratio test should be used to determine if the covariance between parameters is necessary.

A residual error model should, by necessity, be part of the basic PM model. It is useful to start with a combination of additive and proportional error models. If the data does not support either of the error models, the estimate of one of the errors would tend toward zero. As a note of caution, if the base model has not been optimized, especially the structural model component, an initial estimate of an infinitely small value for the additive component of the residual error model may lead to an erroneous elimination of that component of the error model. This should be avoided. It is important to let the nature of the data determine the type of error model to be used. For instance, radioactive decay may be better characterized with a power error model.

5. *Optimize the Random Effects Models.* The random effects models—models for intersubject variability and residual error—should be optimized once the structural model has been optimized. This might mean including interoccasion variability if the data supports it.

6. *Simplest Base Model Is Backbone.* The simplest base model that characterizes the underlying patterns in the data should form the backbone for developing a population model. The principle of simplicity stipulates that models with the minimum number of parameters should be used. This is called the *parsimony principle*. The model with the smallest number of parameters that describe the data well is the most parsimonious model.

7. *Goodness-of-Fit.* It is implied in steps 2 to 6 above that diagnostic plots (e.g., weighted residual versus time, weighted residual versus predicted observations, population observed versus predicted concentrations, individual observed versus predicted concentrations) and a test statistic such as the likelihood ratio test would be used in arriving at the base model (see Section 8.6.1.1 for goodness of fit). Once the base model (with optimized structural and variance models) has been obtained, the next step in the PM model identification process is the development of the population model.

8.4.2 Population Model Development

There are several approaches to population model development that have been discussed in the literature (7, 9, 15–17). The traditional approach has been to make scatterplots of weighted residuals versus covariates and look at trends in the plot to infer some sort of relationship. The covariates identified with the scatterplots are then tested against each of the parameters in a population model, one covariate at a time. Covariates identified are used to create a full model and the final irreducible, given the data, is obtained by backward elimination. The drawback of this approach is that it is only valid for covariates that act independently on the pharmacokinetic (PK) or pharmacokinetic/pharmacodynamic (PK/PD) parameters, and the understanding of the dimensionality of the covariate data is not taken into account.

Maitre et al. (15) proposed an improvement on the traditional approach. The approach consists of using individual Bayesian posthoc PK or PK/PD parameters from a population modeling software such as NONMEM and plotting these parameter estimates against covariates to look for any possible model parameter covariate relationship. The individual model parameter estimates are obtained using a base model—a model without covariates. The covariates are in turn tested to determine individual significant covariate predictors, which are in turn used to form a full model. The final irreducible model is obtained by backward elimination. The drawback for this approach is the same as that for the traditional approach.

A third approach proposed by Mandema et al. (16) was an improvement on the Maitre et al. (15) approach. The first step is similar to that proposed by Maitre et al. (15), but in the second step individual PK/PD parameters are regressed against covariates using generalized additive modeling (GAM). In the final step, NONMEM is used to optimize and finalize the population model. The approach does not discuss how a reduction in the dimensionality of the covariate vector should be handled.

Kowalski and Hutmacher (17) have proposed using the Wald approximation to the likelihood ratio test in conjunction with Schwarz's Bayesian criterion (SBC) to determine the covariates for inclusion in a population PM model. In this approach "all possible models" (with or without each of the covariate parameters in the model) are tested. The process proceeds as follows:

Step 1. Fit the basic model without covariates to the data and estimate the individual Bayesian PM parameters.

Step 2. Plot the individual Bayesian parameters versus covariates (those estimated from the base model in step 1).

Step 3. Assess functional form for the covariates to be entered into the model (e.g., linear versus nonlinear).

Step 4. Fit the full model including all covariates (number of covariates = k).

Step 5. Employ the Wald approximation method (WAM) to screen the 2^k possible submodels to identify the best; the "best" models being the 10–15 models with the highest SBC (these are the approximate SBCs from WAM).

Step 6. Fit the models selected in step 5 in NONMEM to verify concordance between the actual and the approximate SBCs.

Step 7. Plot the empirical Bayes prediction from the final (irreducible) model versus covariates and compare them to the corresponding plots from the base model.

The WAM approach has the advantage of being rapidly executable while testing many competing models and can incorporate time-varying covariates. It has rarely been used, so experience with this approach is limited. WAM generates an approximate SBC and this type of covariate evaluation has been demonstrated to work well when the approximate SBC and the actual SBC are not discordant. However, it is unknown how well WAM selects covariates when the approximate and actual SBCs are discordant and it may not be sensitive to influence data. Also, it is worth noting that the χ^2 test approximation to the likelihood ratio test is generally considered a better approximation than the Wald test.

Ette and Ludden (7) improved on the approaches of Maitre et al. (15) and Mandema et al. (16) by proposing a five-step approach to population modeling:

Step 1. Exploratory data analysis to examine distributions and correlations among covariates.

Step 2. Determination of a basic PK (or PK/PD) model using NONMEM, for example, and obtaining Bayesian individual parameter estimates.

Step 3. Examination of the distributions of parameter estimates.

Step 4. Exploratory modeling with modern regression techniques such as GAM and TBM for the initial selection of covariates and revealing structure in the data.

Step 5. Final NONMEM modeling to determine the population model with the evaluation of the parameter estimates.

The reduction of the dimensionality of the covariate vector by eliminating redundant covariates—taking into account colinearity between covariates—before performing the GAM step is taken into account with this approach.

Ette (9) introduced the concept of model stability that allows the pharmacometrician to ensure that the covariates retained in the final irreducible model are those supported by the data. The steps in this process are as follows:

Step 1. Determine a basic PM model without covariates in the model.

Step 2. Generate 100 bootstrap data sets.

Step 3. Apply the basic model to each of the 100 bootstrap data sets and determine the individual Bayesian PM parameters.

Step 4. Apply GAM to each of the sets of individual Bayesian PM parameters. Here the PM parameter is the dependent variable in the GAM and the covariates are the independent variables. Here set $\alpha = 0.05$ with a selection criterion cutoff value of 50%.

Step 5. Those variables not attaining the cutoff value are removed from further consideration for inclusion in the model.

Step 6. With the appropriate pharmacostatistical models, population model building is performed using covariates retained in step 5 with the covariate selection level set at $\alpha = 0.005$. The backward elimination for covariate selection in applied to each of the 100 bootstrap samples. The covariates found to be important in explaining the variablilty in the parameter of interest are used to build the final population PM model.

Step 7. The covariate-population model is then applied to the original data to obtain the parameter estimates for the drug.

The stability testing approach also permits hypothesis generation in that a covariate identified in the stability testing step but not retained in the final irreducible model may be further investigated in a future study to establish its significance. Model stability assessment should be included in step 4 from the immediately above method to implement an approach to model development that results in developing a model in which substantive errors are most likely absent.

Furthermore, Ette et al. (4) refined the approaches discussed in the preceding two paragraphs and consolidated them into the PM knowledge discovery approach, which encompasses and exceeds population modeling. The PM knowledge discovery approach (see Chapter 14) to population modeling and extracting hidden knowledge from a clinical trial data set is recommended as the best approach for making the most of a clinical trial data set.

The methods reviewed above address primarily hierarchical models but an issue often arises concerning competing nonhierarchical models. That is, which model is the preferred? These models are most often not independent. However, a test statistic can be used to discriminate between models, which is the difference of the minimized objective functions (log-likelihood differences, LLDs) for the two nonhierarchical models (18). In the next section the approach for obtaining the test statistic for comparing the two nonhierarchical models (18) is described.

8.4.3   Comparison of Nonhierarchical Models

Consider a model,

$$C_{ij} = f(\phi_{ij}, x_{ij})\varepsilon_{ij}, \quad i = 1, \ldots, M, j = 1, \ldots, n_i \qquad (8.1)$$

where C_{ij} is the jth concentration in the ith subject, f is a nonlinear function of a subject-specific parameter vector ϕ_{ij} and the predictor vector x_{ij}, ε_{ij} is a normally distributed noise term with zero mean and variance σ^2, M is the total number of subjects, and n_i is the number of concentrations in the ith cluster. The subject-specific parameter vector is modeled as

$$\phi_{ij} = \phi\exp\eta_i, \quad \eta_i \sim N(0, \omega^2) \qquad (8.2)$$

where ϕ is a p-dimensional vector of fixed population parameters, and η_i is a q-dimensional random effect vector (intersubject variability) associated with the ith cluster (not varying with j).

Let M_1 be a model containing p_1 parameters and let M_2 be a model containing p_2 parameters, which are supposed to be a subset of the parameters of model M_1, thus $p_1 > p_2$. The likelihood ratio statistic $LR(M_2,M_1) = -2\log[l(M_2)/l(M_1)]$, where $l(M_1)$ and $l(M_2)$ are maximized likelihood functions of models M_1 and M_2, respectively, and would follow a central χ^2 distribution with $p_1 > p_2$ degrees of freedom under the null hypothesis that additional parameters contained in the model M_1 are all zero.

$LR(M_2,M_1)$ would follow an asymptotically noncentral χ^2 distribution with some noncentrality parameter δ under the alternative that at least one of the additional parameters is nonzero. Thus, the null hypothesis could also be expressed as $\delta = 0$.

Now consider two models X and Y, which have a certain set of p parameters in common, but in addition X contains p_X parameters that are disjoint to the p_Y parameters additionally contained in Y. Let the model characterized by the p common parameters be denoted by XY.

$LR(XY,X)$ will follow asymptotically a noncentral χ^2 distribution with some noncentrality parameter δ_X and p_X degrees of freedom if not all p_X parameters inserted into model Y are zero. Correspondingly, if not all p_Y parameters inserted into model Y are zero, $LR(XY,Y)$ will follow asymptotically a noncentral χ^2 with

some noncentrality parameter δ_Y and p_Y degrees of freedom. Consider the case that the number of additional parameters in either model is the same, that is, $p_X = p_Y$. By testing the hypothesis of $\delta_X = \delta_Y$ against the alternative $\delta_X \neq \delta_Y$, the question of whether improvements in fit (over model XY) by models X or Y are quantitatively different can be addressed. Interpretation in the form of a difference in the non-centrality parameters is ambiguous when $p_X \neq p_Y$.

Note that the two likelihood ratio statistics $LR(XY,X)$ and $LR(XY,Y)$ are not independent. Therefore, testing whether two nonhierarchical models with equal degrees of freedom fit the data equally well is reduced to testing whether the non-centrality parameters of two independent χ^2 distributions with equal degrees of freedom are identical.

The maximum likelihood estimator for the noncentrality parameter $\hat{\delta}_X$ or $\hat{\delta}_Y$, δ_X or δ_Y is a monotone function f of $LR(XY,X)$ or $LR(XY,Y)$. Letting F be the full model, it follows that

$$\hat{\delta} = f[LR(XY, X)] = f[LR(XY, F) - LR(X, F)] \qquad (8.3)$$

and

$$\hat{\delta} = f[LR(XY, Y)] = f[LR(XY, F) - LR(Y, F)] \qquad (8.4)$$

Therefore, $\hat{\delta}_X = \hat{\delta}_Y$ if and only if $LR(X,F) = LR(Y,F)$. Testing this equality can be done by testing the equality of $LR(X,F)$ and $LR(Y,F)$. The latter likelihood ratio statistics are the objective functions (i.e., -2 log-likelihood of the data) of nonlinear mixed effects models. As a test statistic, the difference of objective functions (log-likelihood difference, LLD) of two nonhierarchical models can therefore be used.

An estimate of the sample distribution of this test statistic under the null hypothesis has to be derived to perform a test of the form described above. This can be achieved by using the bootstrap to obtain the sample distribution of the differences of the objective function given the observation. For this method bootstrap data sets are constructed, and for each bootstrap data set the parameters are estimated and the objective functions are reported for each of the competing models. The confidence interval for the differences of the objective functions is calculated and if this interval does not include 0 then the null hypothesis that the models are equal would be rejected. The percentile method for computing the bootstrap confidence interval as described by Efron (19) is used, and 1000 bootstrap replicates are required for this.

In most cases the choice between two competing nonheirarchical models boils down to choosing the model with a more stable formulation. This is because an examination of the diagnostic plots may not yield any difference in how the models characterize the data, and just choosing a model with a lower objective function may not necessarily indicate that it is a better model (e.g., see Ref. 18).

8.5 PARAMETER IDENTIFIABILITY

Identification described above is a parameter estimation problem. The identifiability problem is a more circumscribed problem. It deals with the following question: Given a model of the system under investigation and specific input–output

experiments (models), would the parameters of the model be uniquely determined if the data were error free? The identifiability problem is strictly a mathematical and an a prori problem since it is concerned with the theoretical existence of unique solutions. Thus, what is encountered sometimes with PM models is the issue of model identifiability. One is not just interested in knowing whether or not a model is identifiable given a data set from an experiment; even if it is not, one is interested in knowing which parameters are identifiable. This is because one may only need to estimate some parameters, as in the case of a minimal model, to test a hypothesis. In addition, a set of parameters may be identifiable, but interactions between parameters, as measured by correlations, for instance, may be such as to make numerical estimation of individual parameter values difficult. Thus, a parameter can be identifiable but poorly estimable for a given trial data set (e.g., see Ref. 20 for a detailed review). Thus, it is necessary to design population PK/PD studies to optimize pertinent parameter estimability. Informative sampling design should therefore be used to minimize the errors of estimation. Identifiability, estimability, and informative sampling design (see Chapter 12) should go in concert as linked steps in parameter estimation.

8.6 APPROACHES TO MODEL EVALUATION

8.6.1 Model Evaluation

There are three elements to model evaluation: (a) assessing the model for goodness-of-fit, (b) checking the model for reliability, and (c) checking the model for stability. The question to be addressed here is not "Is the model true?" but "Are the structure and form of the model without significant error?" For example, a question here would be: "Are the data best described by a one- or a two-compartment deterministic model?" or "Which covariates (creatinine clearance, weight, gender) provide an improved fit of the data?" When addressing the form of the model, we are addressing the issue of the configuration of the relationship between a covariate and a parameter. The form could be linear, sigmoidal, or a sine wave relationship between the parameter of interest and a covariate.

8.6.1.1 Goodness-of-Fit

Goodness-of-fit assessments require diagnostic plots, such as the observed dependent variable (ODV) versus the predicted dependent variable (PDV); residuals versus PDV; weighted residuals versus PDV; weighted residuals versus time; and residuals versus covariates to examine for any type of systematic error. For the plot of ODV versus PDV, some prefer to plot PDV on the *x*-axis and ODV on the *y*-axis. Those who hold this view have argued that the interest is in the variability of the ODV about the PDV and the sources of variability (assay error, etc.) are in the ODV not the PDV. However, another school of thought prefers to plot ODV on the *x*-axis because that is what is measured and PDV is predicted and changes with the model. The line of identity (intercept 0 and slope 1) should run through the center of this plotted data. The plot of residuals and weighted residuals should be scattered evenly above and below the 0 reference line (intercept 0 and slope 0). Figure 8.2 presents a plot of ODV versus PDV with the line of identity noted to be
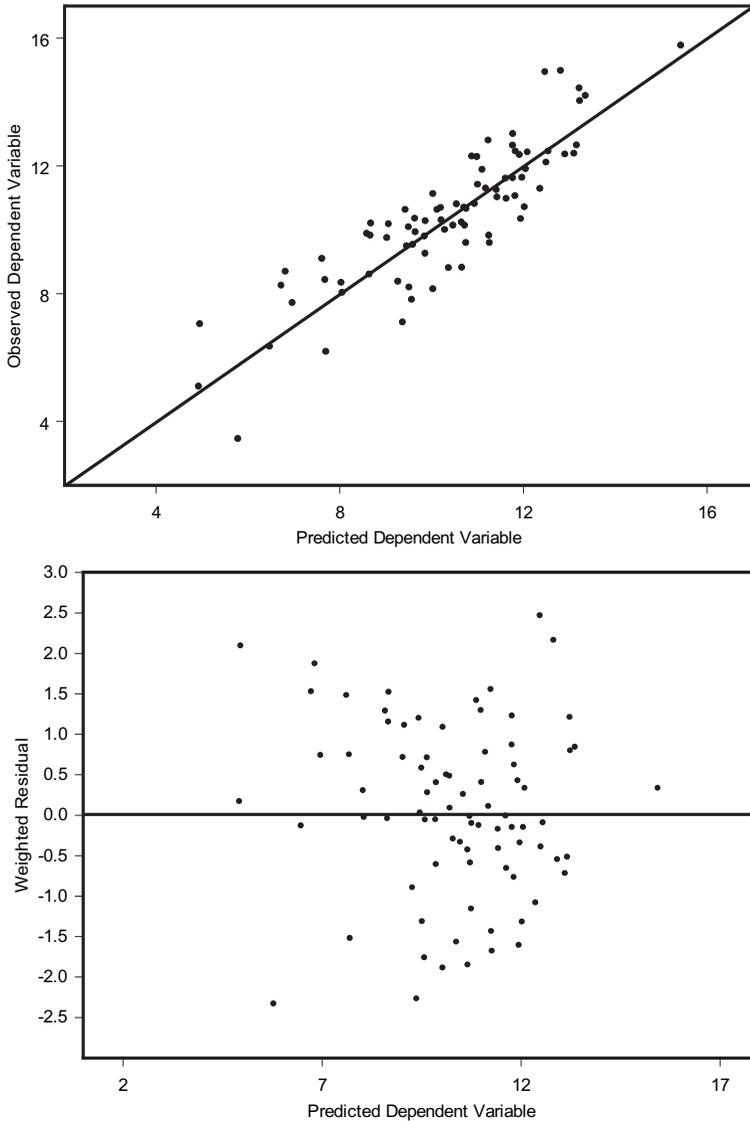
FIGURE 8.2 Plot of observed versus predicted dependent variable (upper panel) and weighted residual versus predicted dependent variable (lower panel).

running through the center of the data and a plot of weighted residuals (WRES) versus PDV with the 0 reference line running through the center of the data. PM models should be without systematic error. A progressive change in variance with such a plot would suggest heteroscedasticity—the band of the residuals would have a nonuniform width. A curved plot or runs in the residuals (like a sine wave) would tend to suggest that the model is inadequate. Heteroscedasticity can be modeled using a multiplicative error model. A combination error model can be used to account for both homoscedastic and heteroscedastic residual errors. Examining a normal scores plot of weighted residuals is useful as a check on the assumption that

the residual errors are normally distributed. It is important to let the data dictate the error model of choice. See, for instance, Refs. 21–23 for more information on error models for PM analysis.

Of importance here is the recent work of Hooker and Karlsson (24) with respect to the plots of the WRES versus PDV when employing NONMEM for model development. They noted that the first-order conditional estimation (FOCE) method is often used to develop population PK and PD models and that WRES are an important diagnostic during population model development. However, even when FOCE is employed to estimate the population parameters, the WRES is computed using the first-order (FO) approximation in NONMEM. It is not clear what statistical properties the WRES should have when using the FOCE approximation. They propose using a new diagnostic tool, the conditional WRES (CWRES), when the FOCE approximation is used. The CWRES is the generalized weighted residual under the FOCE approximation, given the "right model" and adequate data. Statistically the CWRES should be $N(0, 1)$ when FOCE is used. They went on to demonstrate the utility of the CWRES in three examples of correctly specified and misspecified models. For this limited test the CWRES seemed to perform well as a "goodness-of-fit" tool. CWRES may not be used for FOCE with interaction, as its usefulness is yet to be proved.

8.6.1.2 Model Reliability

Reliability of the model requires that the model be assessed for the uncertainty of parameters and random effects. We are interested in the standard errors of estimated parameters and random effects in the model. The uncertainty should be small: for parameters, uncertainty should be less than 25% of the relative standard error and for random effects, it should be less than 35% of the relative standard error (25).

Furthermore, when alternative approaches are applied in computing parameter estimates, the question to be addressed here is: Do these other approaches yield similar parameter and random effects estimates and conclusions? An example of addressing this second point would be estimating the parameters of a population pharmacokinetic (PPK) model by the standard maximum likelihood approach and then confirming the estimates by either constructing the profile likelihood plot (i.e., mapping the objective function), using the bootstrap (4, 9) to estimate 95% confidence intervals, or the jackknife method (7, 26, 27) and bootstrap to estimate standard errors of the estimate (4, 9). When the relative standard errors are small and alternative approaches produce similar results, then we conclude the model is reliable.

8.6.1.3 Model Stability

Model stability addresses the question of how resistant the model is to change. The most direct way to answer this question is to assess whether other plausible or probable data change the model structure or form. The biometrical method that can address stability is the bootstrap. Ette has demonstrated how the bootstrap can be employed to check for model stability by generating other plausible data and determining if the model structure is unchanged for the majority of these bootstrap generated data sets (4, 9). If the model structure or form is not changed as a result of this process, then the model is declared to be stable.

8.7   MODEL VALIDATION

When a model is validated it does not mean that it is considered to be true, which is consistent with Box's previously stated dictum. Validation is most often defined as the evaluation of the predictability of the model developed (i.e., the model structure and form together with the model parameter estimates) and estimated from a learning or index data set when applied to a validation (test) data set not used for model building and parameter estimation. Thus, for validation, we are concerned with the predictive performance of a PM model (28). This addresses the issue of transportability of the PM model. That is, ascertaining whether predicted values from a developed PM model are likely to accurately predict responses in future subjects not used to develop the model (1). There are two broad categories of model validation, external and internal, and these will be discussed next.

8.7.1   External Validation

External validation is the most stringent type of validation. This type of validation can be executed when both input data (index population) to estimate and develop the model and output data (test population) on which the model can be tested exist. It is the application of the developed model to a new data set (validation data set) from another study (28). When a model is validated externally, it provides the strongest evidence for transportability. There are several approaches to internal validation, some of which have been proved to have excessive type I error (6) and various methods are reviewed below.

8.7.2   Internal Validation

Very often a test population of data is not available or would be prohibitively expensive to obtain. When a test population of data is not possible to obtain, internal validation must be considered. The methods of internal PM model validation include data splitting, resampling techniques (cross-validation and bootstrapping) (9, 26–30), and the posterior predictive check (PPC) (31–33). Of note, the jackknife is not considered a model validation technique. The jackknife technique may only be used to correct for bias in parameter estimates, and for the computation of the uncertainty associated with parameter estimation. Cross-validation, bootstrapping, and the posterior predictive check are addressed in detail in Chapter 15.

Data splitting has often been used as an approach to PM model validation. With this approach, the data are randomly divided into an index population and a test population. First the model is estimated from the index population. Then the model is fixed and predictions are made into the test population. If the model is validated, the data are often recombined and the model is reestimated from the combined data. The disadvantage of data splitting is that the predictive accuracy of the model is a function of the sample size resulting from the splitting. Of concern for the data splitting approach to validation is how the data should be split and what proportion should be assigned to the index population and to the test population. Random splitting leads to data sets that have the same variation, other than for chance, and is therefore a weak procedure (34). To maximize the predictive accuracy of data

splitting, alternative approaches have been proposed that employ the entire sample for both model development and assessment (6, 9, 28, 30–33).

The resampling approaches of cross-validation (CV) and bootstrapping do not have the drawback of data splitting in that all available data are used for model development so that the model provides an adequate description of the information contained in the gathered data. Cross-validation and bootstrapping are addressed in Chapter 15. One problem with CV deserves attention. Repeated CV has been demonstrated to be inconsistent: if one validates a model by CV and then randomly shuffles the data, after shuffling, the model may not be validated.

A recently proposed method, the posterior predictive check (PPC), may prove useful in determining whether important clinical features of present and future data sets are faithfully reproduced (31–33). The PPC is addressed in Chapter 15 and will not be discussed further here.

8.7.3   Metrics Applied to Model Validation

Several approaches to quantifying dependent variable predictability have been proposed. The metric should be chosen because of its usefulness and one should be careful not to violate any of the underlying assumptions of the approach. There are instances when one of several metrics could be applied. The use of graphical displays, prediction error, standardized prediction error, and prediction through parameters have been summarized elsewhere and will not be addressed directly here (1, 28). A method proposed by Kleijnen (6) will be reviewed herein.

8.7.3.1   Metrics Applied to Model Validation
A common approach to model validation is to predict the dependent variable from the model in a test data set. The most commonly applied metric is the prediction error of the dependent variable, which is simply

$$PE_{ij} = ODV_{ij} - PDV_{ij} \tag{8.5}$$

where PE_{ij} is the jth prediction error in the ith individual, ODV_{ij} is the jth observed dependent variable in the ith individual, and PDV_{ij} is the jth predicted dependent variable in the ith individual. From the indivual PE_{ij}, the mean of the PE_{ij} can be calculated and this is the mean prediction error (MPE). The 95% confidence intervals (CIs) are constructed around the MPE from

$$CI = MPE \pm 2SE_{MPE} \tag{8.6}$$

where SE_{MPE} is the standard error of the MPE. If this CI contains 0, then the model is said to have adequate predictability and is without significant error.

A problem with this approach to validation is that most often the test data set has more than one observation per individual. These observations of the dependent variable within the same individual are not independent, which is an underlying assumption of this statistic approach. That is, this method is appropriate when only one sample is available per subject (28). Several approaches have been proposed to deal with this problem.

8.7.3.2   Dealing with Replicate Nonindependent Observations Within an Individual

Standardized Prediction Error   The standardized prediction error (SPE) takes into account the variability and correlation of observations within an individual. The SPE_{ij} is the ith standardized prediction error in the jth individual. The SPE_{ij} is the PE_{ij}/SD_{ij}, where SD_{ij} is the standard deviation of the PDV_{ij}.

The mean of the SPE is now calculated with its SE and the mean SPE ($MSPE$) with its CI is constructed. Again this CI should include 0 and the standard deviation of the SPE should include 1. The above methods may be overly conservative as uncertainty in parameters is not taken into account, resulting in an appropriate model being rejected or declared to have substantive error (28).

Plotting of Weighted Residuals (WRES)   The WRES are the residuals (ODV_{ij} – PDV_{ij}) that have been normalized by their standard deviations. These WRES are nearly independent even within the same individual. Thus, when one views the data from many individuals, the correlation that one would expect to see from the several measurements within a single individual should not be seen when observing the WRES (35). For an appropriate model the mean of the WRES should be scattered evenly about zero when plotted against a variable such as subject ID number. Figure 8.3 shows this type of plot. Note that upon observation the prediction from method III seems to have the least bias.

Estimating the Mean Prediction Error with Two Random Effects   Another approach to estimating the mean prediction error that accounts for multiple observations in the same individual has recently been proposed. Here the CI is constructed under the statistical model

$$PE_{ij} = MPE + \eta_i + \varepsilon_{ij} \tag{8.7}$$

where η_i is a random effect representing between-subject variability and is a persistent within-individual shift from the MPE and ε_{ij} is a random effect representing the residual variability (3).

8.7.3.3   Kleijnen Validation Method
Kleijnen (6) has proposed a novel approach to model validation. One computes the PEs as

$$PE_{ij} = ODV_{ij} - PDV_{ij} \tag{8.8}$$

In a further step the ODVs are added to the PDVs, where

$$SUM_{ij} = ODV_{ij} + PDV_{ij} \tag{8.9}$$

so that now for each PE_i there is a SUM_i. Next, one fits the regression

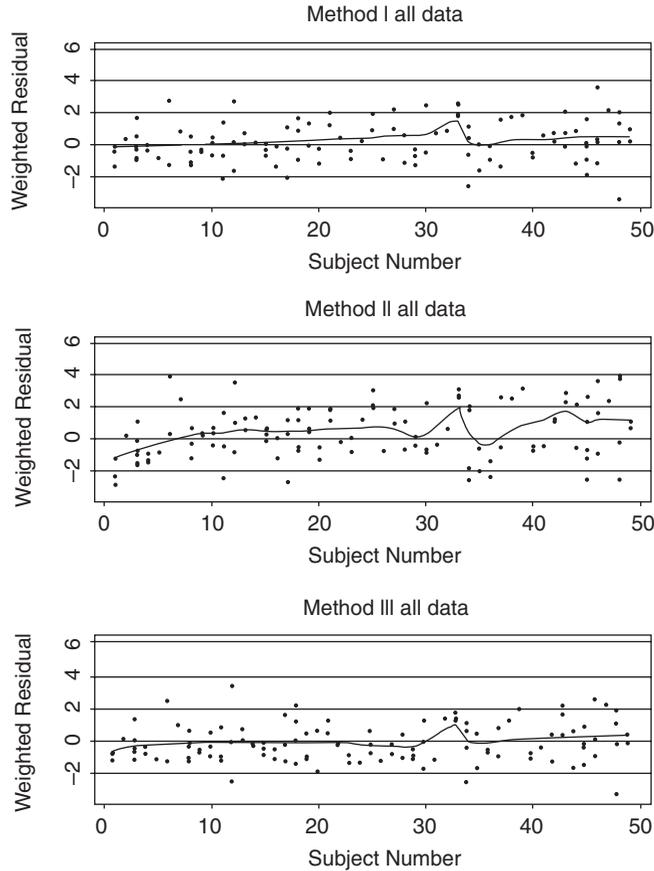$$PE_{ij} = \beta_0 + \beta_1 \bullet SUM_{ij} \tag{8.10}$$

FIGURE 8.3 Plot of weighted residuals versus a random variable, in this case patient ID number (used with permission (Ref. 35)).

to these pairs. The hypothesis is H_0: $\beta_0 = 0$ and $\beta_1 = 0$. This approach assumes normal distributions and independence of all the plotted pairs. This method has not been investigated for population models.

8.7.4 Summary of Prediction Metrics and Methods

Several approaches to model validation have been discussed. No single approach can be applied to all models, and for some models any of several approaches would suffice. In the next section, we present an example of the appropriateness of a predictive model.

8.8 APPLICATION EXAMPLE

8.8.1 The Problem

An oncology agent was being developed and a dosing strategy for a Phase 3 study needs to be proposed. This is a predictive model (1) and therefore needs to be validated. The concept of descriptive versus predictive models has been presented in detail elsewhere (1).

8.8.2 Methods

The drug being developed was administered by rapid intravenous infusion every 6 hours for 48 hours at 3 week intervals. Prior to the next dose the blood work was done to determine appropriateness of administering the next dose cycle. It is proposed that this procedure be continued in Phase 3 and probably eventually ending up on the label.

Prior to the Phase 2 study, simulations were done to determine if an external validation set of data would be an adequate approach to model validation. Of importance here was whether this approach to validation would accept an appropriate model and reject an inappropriate model. A set of simulated data were generated with 150 subjects each with four concentrations. One hundred of these were assigned to the index population completely at random and the other 50 were assigned to the test population completely at random. For the template, the patients' weights (mean of 65 kg; lognormal distribution) and sex were generated in ZRandom software (ZRandom, Sydney, Australia). Five simulated data sets were generated where the index and test population pharmacokinetic (PPK) parameters were the same; five more data sets were generated where the mean clearance was 20% below the test, and five more where the mean clearance was 20% greater than the test. For the simulation, all of these data sets had typical values for clearance and volume specified and also random effects for clearance and volume specified and finally the residual random effect specified. The models estimated from the index data were used to make predictions into the test data. For the 50 data sets each above and below the test clearance values, the models were invalidated by the external validation process in 48 of 50 for the below clearance and 47 of 50 for the above clearance cases. For the 50 data sets where the index model and the test model were the same, all 50 models were validated. On this basis it was determined that prediction into a test data set of 50 subjects would be adequate to validate the model if subjects were assigned to the index and test data completely at random.

In Phase 2 PK data were collected on 150 subjects. One hundred subjects were assigned to the index data and 50 to the test data completely at random and it was determined to collect 400 samples in the index population and 200 samples in the test population. The nominal sample collection times were 0.5–2, 4–6, 48–50, and 52–54 hours after the first dose. Of the 400 scheduled samples in the index population, 375 were obtained and of the 200 scheduled samples in the test population 186 were obtained. The PK knowledge discovery process (see Chapter 14) was used to estimate the PPK model. The final irreducible was

$$Clearance = 1.02 * Weight^{0.589}$$

$$Volume = 2.16 * Weight$$

The coefficient of variation for clearance was 40.2% and for volume it was 38.8%, and residual variability was 14.4%.

This irreducible PPK model was fixed and predictions were made into the test data. The mean prediction error was −0.020 and the standard error of the mean prediction error was 0.022. Therefore, the model was considered to be validated.

The validated model from Phase 2 was used to simulate the range of expected concentrations for Phase 3 at a dose of 10 mg/kg. A template was created for 1000 patients with weight being generated from the ZRandom number assuming a
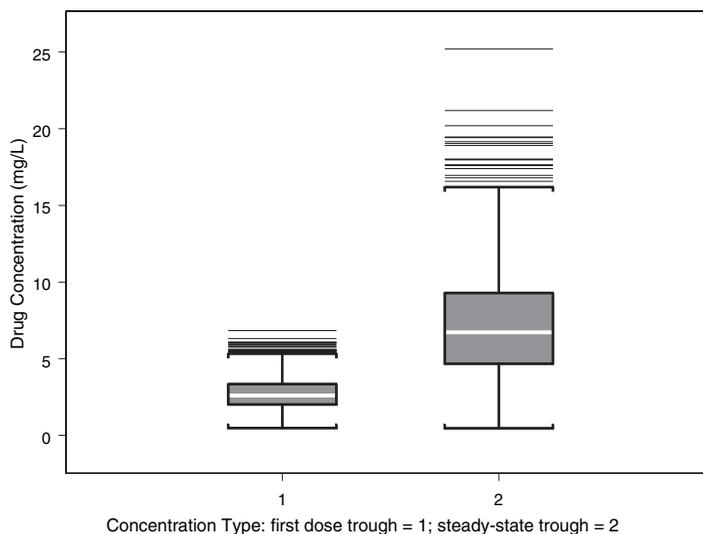
FIGURE 8.4 Box and whisker plot of trough concentrations.

lognormal distribution. It had been determined from Phase 2 that optimal responses were most evident in patients with end of dose interval concentrations above 1.0 mg/L. Therefore, of primary interest was what the range of concentrations would be at the end of the first interval and at steady state. These predicted concentrations are presented in the box and whisker plot (Figure 8.4). Of the 1000 concentrations at the end of the first dosing interval, only 11 would be expected to be below 1.0 mg/L and at steady state only 5 would be expected to be below 1.0 mg/L. Therefore, the 10 mg/kg dosing strategy was adopted.

8.9 SUMMARY

The epistemology of model appropriateness is defined, explained, and demonstrated using a predictive model. A clear process for obtaining pharmacometric knowledge is outlined. It involves model appropriateness, which includes clearly identifying and stating the problem, stating the application or the intended use of the model, evaluating the model, and determining the predictive performance of the model (e.g., via the bootstrap) if the model is to be used for a predictive purpose. A sound epistemological approach should result in a rational approach to model development, evaluation, and validation.

REFERENCES

1. E. Ette, P. Williams, J. Lane, Y. K. Kim, and E. V. Capparelli, The determination of population pharmacokinetic model appropriateness. *J Clin Pharamcol* **43**:2–15 (2003).
2. V. F. Cosson, E. Fuseau, C. Efthymiopoulos, and A. Bye, Mixed effect modeling of sumatriptan pharmacokinetics during drug development. I. Interspecies allometric scaling. *J Pharmacokinet Biopharm* **25**:149–167 (1997).

3. P. J. Williams, J. R. Lane, W. Murray, M. Mergener, and M. Kamigaki, Pharmacokinetics of the digoxin–quinidine interaction via mixed-effect modelling. *Clin Pharmacokinet* **22**:1–10 (1992).

4. E. I. Ette, P. Williams, H. Sun, E. Fadiran, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).

5. R. G. Seargent, Verification, validation, and accreditation of simulation models, in *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick (Eds.). IEEE Standards Office, Los Alamitos, CA, 2000, pp. 647–654.

6. J. P. C. Kleijnen, Validation of models: statistical techniques and data availability, in *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Embhard, D. T. Sturrock, and G. W. Evans (Eds.). IEEE Standards Office, Los Alamitos CA, 1999, pp. 647–654.

7. E. I. Ette and T. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**(12):1845–1855 (1995).

8. J. Mandema, D. Verotta, and L. B. Sheiner, Building population pharmacokinetic–pharmacodynamic models, in *Advanced Pharmacokinetic and Pharmacodynamic Systems Analysis*, D. Z. D'Argenio (Ed.). Plenum Press, New York, 1995, pp. 69–86.

9. E. I. Ette, Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* **37**:486–495 (1997).

10. F. Mentre and M. E. Ebelen, Validation of population pharmacokinetic–pharmacodynamic analyses: review of proposed approaches, in *The Population Approach: Measuring and Managing Variability in Response Concentration and Dose (COST B1)*. Office for Official Publications of the European Communities, Brussels, 1997, pp. 146–160.

11. F. A. Harrell, *Regression Modeling Strategies: With Application to Linear Models, Logistic Regression and Survival Analysis.* Springer-Verlag, New York, 2001, pp. 87–101.

12. G. E. P. Box, Robustness in the strategy of scientific model building, in *Robustness in Statistics*, R. L. Laurer and G. N. Wilkinson (Eds.), Academic Press, New York, 1979, pp. 201–236.

13. R. N. Bergman and C. Cobelli, Minimal modeling partition analysis and the estimation of insulin sensitivity. *Fed Proc* **39**:110–115 (1980).

14. L. B. Sheiner, B. Rosenberg, and V. V. Marathe, Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm* **5**:445–479 (1977).

15. P. O. Maitre, M. Buhrer, D. Thomson, and D. R. Stanski, A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam. *J Pharmacokinet Biopharm* **19**:377–384 (1991).

16. J. W. Mandema, D. Verotta, and L. B. Sheiner, Building population pharmacokinetic–pharmacodynamic models. I. Models for covariate effects. *J Pharmacokinet Biopharm* **20**:511–528 (1992).

17. K. G. Kowalski and M. M. Hutmacher, Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *J Clin Pharmacol* **28**:253–275 (2001).

18. E. I. Ette, Comparing non-hierarchical models: application to nonlinear mixed effects modeling. *Comput Biol Med* **26**:505–512 (1996).

19. B. Efron, Comparing non-nested linear models. *J Am Stat Assoc* **79**:791–803 (1981).

20. C. Cobelli and J. J. DiStefano III, Parameter and structural identifiability concepts. *Am J Physiol* **239**:R7–R24 (1980).

21. M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measures Data*. Chapman & Hall/CRC Press, New York, 1995.

22. M. O. Karlsson, S. L. Beal, and L. B. Sheiner, Three new residual error models for population PK/PD analyses. *J Pharmacokinet Biopharm* **23**:651–672 (1995).

23. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. NONMEM Project Group, University of California, San Francisco, 1999.

24. A. Hooker and M. O. Karlsson, Conditional weighted residuals—a diagnostic to improve population PK/PD model building and evaluation. *Am Assoc Pharmaceut Sci J* **7**(S2): Abstract W5321 (2005).

25. E. I. Ette, H. Sun, and T. M. Ludden, Balanced designs in longitudinal population pharmacokinetic studies. *J Clin Pharmacol* **38**:417–423 (1998).

26. B. Efron, Bootstrap methods: another look at the jackknife. *Ann Stat* **7**:1–26 (1979).

27. B. Efron and G. Gong, A leisurely look at the bootstrap, the jackknife and cross-validation. *Am Statistician* **37**:36–48 (1983).

28. *Guidance on Population Pharmacokinetics for Industry*, Food and Drug Administration, Rockville, MD, 1999.

29. B. Efron, Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* **78**:461–470 (1983).

30. B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

31. A. Gelman, X. L. Meng, and H. Stern, Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**:733–807 (1996).

32. S. B. Duffull, S. Chabaud, P. Nony, L. Chritian, P. Girard, and L. Aarons, A pharmacokinetic simulation model for ivabradine in healthy volunteers. *Eur J Phamaceut Sci* **10**:285–294 (2000).

33. S. B. Duffull and L. Aarons, Development of a sequential linked pharmacokinetic and pharmacodynamic simulation model for ivabradine in healthy volunteers. *Eur J Pharmaceut Sci* **10**:275–284 (2000).

34. R. P. Hirsch, Validation samples. *Biometrics* **47**:1193–1194 (1991).

35. P. J. Williams, J. Lane, E. Capparelli, Y. Kim, and R. Coleman, Comparison of three methods of predicting serum digoxin concentrations. *Pharmacotherapy* **16**:1085–1092 (1996).

Data Imputation

ENE I. ETTE, HUI-MAY CHU, and ALAA AHMAD

9.1 INTRODUCTION

Missing data may bias model parameter estimates, inflate Type I and Type II error rates, and degrade the performance of confidence intervals. Missing values may dramatically reduce statistical power because a loss of data is nearly always accompanied by a loss of information. Pharmacometricians who wish to mitigate these risks must pay close attention to the issue of missing data in the analysis of clinical trials and choose their strategy carefully. Recent computational and theoretical advances, most notably multiple imputation methods, enable the pharmacometrician to use the existing data to generate, or impute, values approximating "real" value, while preserving the uncertainty of the missing values (1).

This chapter provides an overview of imputation, gives a description of incomplete data types, and reviews the standard methods of handling missing data, with a focus on multiple imputation.

9.2 DATA IMPUTATION

The idea of data augmentation arises naturally in missing value problems, as exemplified in the standard ways of filling missing cells in, for instance, balanced two-way tables from clinical trials. Data imputation refers to a scheme of filling in missing data to complete the observed data set to make it easier to analyze. When the data is complete, the quality of the analysis is improved. When a data subset, such as data on a surrogate endpoint or a biomarker, is missing because of sample handling error, data imputation is inevitable if knowledge is to be gained of some aspect of the response surface that is contained in or dependent on the missing data. In such situations, data imputation is an absolute necessity.

9.3   DESCRIPTION OF INCOMPLETE DATA TYPES

Incomplete or missing data is a problem in population pharmacokinetic/pharmaco-dynamic (PK/PD) analysis in that analytic power can be compromised and biased estimates can become the outcome of a time-consuming analysis. When covariates are missing from some subjects, deleting those subjects and analyzing data from only subjects with complete information can reduce analytic power unless the covariate data are missing completely at random. Discarding subjects can bias a study severely. The price of "missingness" becomes even more expensive if the response variable data is missing in a region of the response surface where the information could be crucial.

Identifying whether data is missing completely at random (MCAR) or missing at random (MAR) is key to determining how to handle missing data. Data is said to be missing *completely* at random if the probability of missing data on one variable is not related to the value of that variable or to other variables in the data set (2). If, for instance, failure of a subject to return for a clinic visit for disease evaluation, and therefore measurement of a response variable such as blood pressure (BP), was due to neither her disease status nor her sex, it would be concluded that the values for the response variable were MCAR. On the other hand, if discovery were to be made that males tended to skip clinic appointments, the response variable would not be MCAR. Whether data are MCAR can be verified partly by comparing data from subjects with incomplete response variable measurements against those with complete measurements on all other variables. It would be nearly impossible, however, to determine whether the probability of missing data was a result of the value of the variable itself.

A less restrictive notion than MCAR is data missing at random (MAR). MAR occurs when the probability of the missing value is not dependent on the value itself, but may depend on the values (through correlation, not conditionally) of other variables in the data set (2). This assumption, called "ignorable missingness," also underlies random effects models for longitudinal data with unbalanced patterns and hence missing data (3, 4). If in the above case, for example, the missingness of the variable BP was not dependent on BP, it would be concluded that BP values are MAR, regardless of whether other variables are associated with the missing BP values. By contrast, if high or low body weight individuals tended to miss clinic visits for BP measurements, missing BP is very dependent on weight. It would be concluded BP data values are not MAR.

9.4   APPROACHES FOR HANDLING DATA INCOMPLETENESS

Traditionally, incomplete (missing) data have been handled by deletion from analysis of cases that contain missing values (single imputation) and most recently by using multiple imputation techniques. In this section single imputation techniques are discussed. Multiple imputation and the paradigm for multiple imputation are discussed in separate sections because of the broad scope of these topics.

9.4.1 Casewise Deletion

With casewise deletion data records or cases with missing data are simply discarded, restricting the analysis to those records or cases with a full complement of values. The shortcomings of various case-deletion strategies have been well documented (5, 6). If the data are MCAR (i.e., the probabilities of response variable being measured do not depend on any data values observed or missing), deletion yields unbiased parameter estimates but larger standard errors because of reduced sample size. Casewise deletion, however, can lead to misleading results if a large proportion of the data is discarded. This is also problematic if the data are not MCAR. If the data are MAR, but not completely at random, casewise deletion may lead to biased estimates, that is, regression coefficients that are erroneously too large or too small (7). Casewise deletion assumes that deleted cases are a random subsample of the data set (8), which of course is erroneous. Analytic power will be severely compromised if a large proportion of the data is missing.

9.4.2 Maximum Likelihood Value and Bayesian Estimation

Accurate results may be obtained by maximum likelihood (ML) estimation or Bayesian estimation if one is using a formal probability model (e.g., a normal model) and the missing values are MAR when dealing with missing data. Since both ML and Bayesian approaches rely on the complete data likelihood, the function linking the observed and missing data to the model parameters, the probability model is key.

Provided the data missingness mechanism is MAR, the ML and Bayesian approaches are useful for the analysis of incomplete data. They estimate parameters of interest without requiring one to impute the missing data in the data set. The requirement of fairly sophisticated computational methods and model specificity is a disadvantage of these methods. In addition, these methods do not model the missingness mechanism, which is equivalent to assuming that the missing values are MAR, and may yield erroneous results if the missingness mechanism is missing not at random (MNAR).

9.4.3 Single Imputation

Single imputation involves ascribing a value to a missing data cell based on other variables or of substituting a reasonable estimate for absent data elements (5). With single imputation, one value is ascribed to the missing value. A mean value, for example, is sometimes used to represent missing data. Imputing the mean eliminates data that may be unique to a particular individual and ascribes the "typical value" to that subject. Imputing the mean may sometimes lead to erroneous statistical inferences (6). Mean imputation decreases variability between individuals and biases correlations with other variables (9). With single imputation techniques variability between imputations is not accounted for because only one value is imputed (8). This results in increasing the sample size, but decreasing the variance by use of a mean value for substitution. There is a problem associated with decreased variances: that is, the estimates are too close to the mean (9). When the rate of missing information

is small (say, <5%), then single imputation inferences for a scalar estimand may be fairly accurate. For joint inferences about multiple parameters, however, even small rates of missing information may seriously impair a single imputation procedure.

Hot deck is another type of single imputation. This procedure matches individuals with missing data with those having similar values in a set of other variables and imputes the known value into the missing data cell. The flaw with hot deck imputation is its treatment of imputed data with certainty, thus perhaps grossly underestimating variability (6).

Another type of imputation is (*multiple*) *linear regression*. Predictive equations generate the imputed values using complete-case information. This method may work well when predictors are strong, but here too the absence of sufficient variability causes underestimation of standard errors (5). An alternative and better procedure is *stochastic regression* in which each missing value in the data set is replaced by the predicted value from a regression analysis based on complete cases plus a random residual term (10). The stochastic regression imputation method improves on regression imputation if the regression model is reasonable. Both of these techniques, however, can result in biased parameter and standard error of estimates.

Last observation carried forward (LOCF) is another example of a single imputation method. In a clinical trial, once a patient withdraws from the treatment protocol, it may be difficult or even impossible to continue data collection on the same schedule. In such trials, the study database can suffer from substantial truncation by nonadherence or dropout. In that case, dropout causes a particular pattern of missing data: each patient has a last occasion of measurement, all responses are observed up to that occasion, and there are no responses observed after that occasion. This is called a monotone pattern unit-level missing data (11). This missing data pattern violates the strict "intention-to-treat" (ITT) principle (12): measure all subjects' outcomes regardless of protocol adherence and analyze by treatment assigned no matter what treatment subjects actually received. With LOCF, a missing value at a time t is imputed by the most recent observation of the subject at time $t' < t$. In performing ITT analysis with clinical trial data, LOCF has generally been the imputation approach used to handle monotone pattern unit-level missing data. The LOCF method works best if it is known that the post dropout values remain frozen at the last observed value. This is usually an untenable assumption. When the timing and rate of withdrawal differ among treatments, the interpretation of results obtained by LOCF is problematic (13). This imputation approach may complete data sets for the desired analysis to be performed but the statistical inference derived may be erroneous. If, for instance, placebo subjects in a study have a worsening course of their disease and the dropouts are concentrated in the treatment group, the LOCF method may show an ineffective treatment to be effective by interrupting the deteriorating course after dropout. On the other hand, if placebo subjects have a worsening course and the dropouts are concentrated in this group, then the LOCF method may show an effective treatment to be ineffective by interrupting the deteriorating course after dropout of subjects in the placebo group. Moreover, if the treatment has a carry-over effect after dropout and there are more dropouts in the early period of follow-up in the treatment group (possibly due to side effects), then the LOCF method underestimates the treatment effect.

Generally, single imputation can easily be implemented and thus allows the application of the standard complete-data method of analysis. However, the under-

estimation of uncertainty with single imputation can be a major problem. It treats imputed values as if they were true and thus overestimates precision (5). As previously discussed, variability decreases, affecting the plausibility of parameter estimates and associated uncertainty terms.

9.5   MULTIPLE IMPUTATIONS

9.5.1   Concepts, Assumptions, and Constraints

Multiple imputation (MI), developed by Rubin (2, 6, 14), is a predictive approach to handling missing data in multivariate analyses. It blends both classical and Bayesian statistical techniques and relies on specific iterative algorithms to create several imputations. MI rectifies the major disadvantages of single imputation by replacing each missing value with a vector composed of $M \geq 2$ possible values (usually between 2 to 10 possible values, but commonly 5) to accurately reflect uncertainty and to preserve important data relationships and aspects of the data distribution. It requires that the analyst specifies an imputation model, imputes several data sets, analyzes them separately, and then combines the results. MI yields a single set of test statistics, parameter estimates, and standard errors.

The validity of the method hinges on how the imputations are generated. It is not possible to obtain valid inferences if imputations are created arbitrarily. On average, the imputation should give reasonable predictions for the missing data, and variability among them should reflect an appropriate degree of uncertainty. Rubin (6) provides technical conditions under which a repeated-imputation method leads to frequency-valid answers. An imputation method that satisfies these conditions is said to be "proper" (6). Stated simply, procedures for imputation, whether based on explicit (parametric) or implicit (nonparametric) models or based on ignorable or nonignorable models, that incorporate appropriate variability among repetitions within a model are called "proper." A variety of proper imputation methods based on both explicit and implicit models, including a fully normal model, the Bayesian bootstrap, and the approximate Bayesian bootstrap (ABB), have been studied by Rubin (15). An imputation model must preserve all important associations among variables in the data set, including interactions if they will be part of the final analysis. Also, the dependent variable must be included in the model to ensure that all relationships between variables are maintained (1). Finally, the algorithm used to generate imputed values must be "correct"; that is, it must accommodate the necessary variables and their associations. Allison (16) illustrated this fact by comparing disparate results of two algorithms for producing multiple imputations. The first algorithm considered only the variables associated with the missingness of the data, and the second included other variables and their associations. Allison's findings clearly support Rubin's (6) contention that good imputation methods use all information related to missing cases.

9.5.2   Advantages and Disadvantages

MI builds on the advantages of single imputation. It allows the use of complete-data analysis methods for data analysis and also includes the data collector's knowledge. In addition, it incorporates random error because it requires random variation in

the imputation process. MI produces improved estimates of standard errors when compared with single imputation methods because repeated estimations are used. It can accommodate any model and any data and does not require specialized software. MI also increases efficiency of parameter estimates because it minimizes standard errors and simulates proper inferences from data (6).

The three disadvantages of MI when compared with other imputation methods are: (a) more effort to create the multiple imputations, (b) more time to run the analyses, and (c) more computer storage space for MI-created data sets (6). These are hardly issues with current development in computer technology. The MI approach is computationally simpler than the ML and Bayesian approaches for most practical situations. Once the imputed data is generated, the data can be analyzed with any data analysis software of choice.

9.6   THE MI PARADIGM

9.6.1   Requirements for Model-Based MI

Multiple imputations are generated by assuming a particular imputation model. Therefore, the success or failure of MI depends on the propriety of the assumed imputation model. Assumptions required in MI are (a) a model for the data values, (b) a prior distribution for parameters of the data model, and (c) the nonresponse mechanism. However, with nonparametric methods of MI, minimal distributional assumptions are required (see Section 9.6.6).

9.6.2   Data Model

Assuming a probability model that relates the complete response (or dependent) data Y (the combination of observed values Y_{obs} and the missing values Y_{mis}) to a set of parameters is the first and most important step to obtaining multiple imputations. With the probability model and the prior distribution on parameters (see Section 9.6.3), a predictive distribution $P(Y_{mis}|Y_{obs})$ for the missing values conditional on the observed values is found, and the imputations are then generated from the predictive distribution.

The assumed model needs to incorporate all the knowledge one has about the process that generated the data. The multivariate normal assumption is the most convenient model for continuous variables. Then the model is manageable computationally is a key advantage. The multivariate normal model gives quite acceptable results even when the variables are binary or categorical, with the imputations performed using a normal model and then the imputed values are rounded off to the nearest category (1). A variable that is not normally distributed may be transformed to a normal variable and the imputed values are then transformed back to the original scale. Others have used a log-linear model for categorical variables, a mixture of a log-linear and a multivariate normal model for mixed continuous and categorical data sets, and a hierarchical linear model (17).

9.6.3   Prior Distribution

Model-based MI is usually performed using a Bayesian statistical approach. Thus, there is a need for specifying a prior distribution on the parameters to carry out

the analyses. The prior distribution and the complete-data model provide the predictive distribution $P(Y_{\text{mis}}|Y_{\text{obs}})$ for the missing data conditional on the observed values from which the imputations can be generated. For convenience, however, noninformative prior distributions are used to do MI. The subjectivity involved in the choice of prior distributions have, at times, led to a criticism of the Bayesian methods. Prior distributions hardly matter for many data analyses because with even moderately large sample sizes any reasonable prior distribution gives essentially the same results. With a small sample size, doing the analysis under different prior distributions and examining the results for change—a type of sensitivity analysis—is a reasonable check before drawing any conclusions.

9.6.4 Missing-Data Mechanism for MI

MI that is model based assumes that the missing data are MAR. This assumption allows one to use the relationships among the variables evident from the observed data to obtain imputed values for the missing data.

9.6.5 Parametric Bayesian Models

Rubin's suggested Bayesian approach to MI was popularized by Schafer (1), who provided detailed algorithms for creating MIs in different situations. Suppose, in general, that $Y = (y_1, y_2, \ldots, y_n)$, where the first a values $[Y_{\text{obs}} = (y_1, y_2, \ldots, y_a)]$ are actual observed values and the remaining values $[Y_{\text{mis}} = (y_{a+1}, y_2, \ldots, y_n)]$ are missing at random. $Y = (Y_{\text{obs}}, Y_{\text{mis}})$ follows a parametric model $Y \sim P(Y|\theta)$, where θ is the unknown parameter, or a vector of parameters in the multivariate case, that we are ultimately interested in (e.g., mean, variance, or shape that describes the response surface). θ is assumed to have a prior distribution and Y_{mis} is ignorably missing. MIs are Bayesianly proper if they are independent realizations of $P(Y_{\text{mis}}|Y_{\text{obs}})$, the posterior predictive distribution of the missing data under some complete-data model and prior. $P(Y_{\text{mis}}|Y_{\text{obs}})$ may be written

$$P(Y_{\text{mis}}|Y_{\text{obs}}) = \int P(Y_{\text{mis}}|Y_{\text{obs}}, \theta)P(\theta|Y_{\text{obs}})d\theta \qquad (9.1)$$

the conditional predictive distribution of Y_{mis} given θ, averaged over the observed-data posterior of θ. Thus, Bayesianly proper imputations reflect uncertainty about Y_{mis} given the parameters of the complete-data model, as well as uncertainty about the unknown model parameters. The resulting MIs are appropriate under an assumption of ignorability because $P(Y_{\text{mis}}|Y_{\text{obs}})$ does not rely on the pattern of the observed response. Thus, an imputation for Y_{mis} can be described in two steps: first by simulating a random draw of the posterior distribution of the unknown parameter $\theta^* \sim P(\theta|Y_{\text{obs}})$ and followed by a random draw of the missing values from their conditional predictive distribution

$$Y_{mis}^* \sim P(Y_{\text{mis}}|Y_{\text{obs}}, \theta^*) \qquad (9.2)$$

For some cases, the posterior distribution of θ is not straightforward, due to a nonstandard distribution that cannot easily be simulated. Rubin (6) introduced a

few general strategies for approximating draws for the posterior distribution of θ, including large-sample normal approximations and importance resampling. Moreover, a more popular Markov chain Monte Carlo (MCMC), which creates a Markov chain with a desired stationary distribution, provides an ideal method suited to handle missing-data problems (18). Overview of MCMC algorithms—including Gibbs sampling, Metropolis–Hastings algorithm, and the data augmentation algorithm—are provided by Gilks et al. (18) and Tanner and Wong (19).

9.6.5.1 MCMC for MI

The data augmentation algorithm developed by Tanner and Wong (19) is an MCMC method ideally suited to missing-data problems. It is an iterative two-step process in which missing observations are alternatively sampled from their conditional predictive distribution $Y_{\text{mis}}^{(t)} \sim P(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t-1)})$ and then unknown parameters are sampled from a simulated complete-data posterior $\theta^{(t)} \sim P(\theta|Y_{\text{obs}}, Y_{\text{mis}}^t)$. This defines a Markov chain $\{(Y_{\text{mis}}^{(t)}, \theta^{(t)}), t = 1, 2, \ldots\}$, given an initial value $\theta^{(0)}$, which under quite general conditions converges to the stationary distribution $P(Y_{\text{mis}}|Y_{\text{obs}})$. When these steps are executed a large number of times, it eventually produces a draw of θ from its observed data posterior $\theta^* \sim P(\theta|Y_{\text{obs}})$, and a draw from Y_{mis} from $P(Y_{\text{mis}}|Y_{\text{obs}})$, the distribution from which MIs are produced. The second step of data augmentation $\theta^{(t)} \sim P(\theta|Y_{\text{obs}}, Y_{\text{mis}}^{(t)})$ is straightforward in many cases. This step may be intractable in more complicated situations and may be replaced by one or more cycles of another MCMC algorithm that converges to $P(\theta|Y_{\text{obs}}, Y_{\text{mis}}^{(t)})$.

Schafer (1) has described MCMC methods for basic models for continuous, categorical, and mixed multivariate data; and MCMC methods provide a flexible set of tools for creating MIs from parametric models.

9.6.6 Nonparametric Bayesian Methods

Rubin (15) describes a simple method called the approximate Bayesian bootstrap (ABB). This approach makes it possible to generate proper imputation for Y_{mis} with minimal distributional assumptions. To illustrate the ABB approach for MI, consider a collection of n units with the same value of predictor X, where a subjects were observed and $n_{\text{mis}} = n - a$ subjects with missing values. The ABB creates M ignorable repeated imputations from $m = 1, \ldots, M$ as follows: (a) create a new pool of Y_{obs}^* by sampling a values from $Y_{\text{obs}} = (y_1, y_2, \ldots, y_a)$ with replacement, and (b) select a set of n_{mis} possible values from Y_{obs}^*, again with replacement. By drawing n_{mis} missing values from a *possible* sample of Y_{obs}^* values rather than from the Y_{obs} values, the ABB approach generates appropriate between-imputation variability, at least assuming large random samples at X, as demonstrated by Rubin and Schenker (20).

9.6.7 Combining Estimates

After M imputations have been created for a data set, they are then analyzed using standard PK/PD or statistical package. There are now M complete data sets containing the observed values and the imputed values instead of one. The PK/PD analysis must be done M times, once on each complete data set. Across M data

sets the results will vary, reflecting the uncertainty due to missing observations. The M complete-data analyses are combined to create one repeated-imputation inference.

Let $\hat{\Theta}_m$ and U_m, $m = 1, \ldots, M$, be M complete-data estimates and their associated variances for a parameter Θ, calculated from the M data sets completed by repeated imputations under one model for unobserved data. In the linear regression case, for instance, $\Theta = \beta$, $\hat{\Theta}_m$ is the least squares estimate of β and U_m is the standardized residual mean square error. The repeated imputation estimate of Θ is the mean of the complete-data estimates:

$$\bar{\Theta} = \sum_{m=1}^{M} \left(\hat{\Theta}_m / M \right) \tag{9.3}$$

There are two components of the variability associated with this estimate, the average within-imputation variance,

$$\bar{U} = \sum_{m=1}^{M} \left(U_m / M \right) \tag{9.4}$$

and the between-imputation component,

$$B = \sum_{m=1}^{M} \left[\left(\hat{\Theta}_m - \bar{\Theta} \right)^2 / (M - 1) \right] \tag{9.5}$$

The total variability associated with $\bar{\Theta}$ is given by

$$T = \bar{U} + \left(1 + M^{-1} \right) B \tag{9.6}$$

Inference can be made using Θ, T, and a distributional assumption. For example, if Θ is a scalar quantity, the approximate reference distribution for interval estimates and significance tests is a t distribution:

$$\left(\Theta - \bar{\Theta} \right) T^{-1/2} \sim t_v \tag{9.7}$$

where the degrees of freedom, v, are given by

$$v = (M - 1)\left(1 + r^{-1} \right)^2 \tag{9.8}$$

with

$$r = \left(1 + M^{-1} \right) B / \bar{U} \tag{9.9}$$

Thus, a $100(1 - \alpha)\%$ interval estimate for $\bar{\Theta}$ is

$$\bar{\Theta} \pm t_{v, 1-\alpha/2} \sqrt{T} \tag{9.10}$$

The between-subject and within-subject ratio, r, estimates the population quantity $\gamma/(1 - \gamma)$, where γ is the fraction of information about Θ missing due to sample

handling error, unobserved data, and so on. The same approach is used for combining parameter estimates in the nonlinear mixed effects regression scenario.

9.6.8   Multiple Imputation for Truncated Data

Left-censored data are characteristic of many bioassays due to the inherent limitation of the presence of a lower limit of detection and quantification. An ad hoc approach to dealing with the left-censored values is to replace them with the limit of quantification (LOQ) or LOQ/2 values. Alternatively, one can borrow information from other variables related to the missing values and use MI to estimate the left-censored data. In addition, the left-censored mechanism can be incorporated directly into a parametric model, and a maximum likelihood (ML) approach can be used to estimate the parameters (21).

As previously stated, an important assumption behind MI is the "ignorability of missingness," either MCAR or MAR. When the missingness is not ignorable, as in cases where the dropout pattern is such that dropout patients all have deteriorating symptoms prior to their dropping out, the "ignorability of missingness" assumption is violated. Dropout can also occur as a result of censoring applied to individual measurements or because some of the units are withdrawn from the study prematurely. Another good example of dropout directly related to the measurement process is when ethical considerations may require a patient to be withdrawn from a trial on the basis of his/her observed measurement, for example, blood pressure (BP) is not adequately controlled in a long-term trial of a drug tested for BP reduction. It is important, therefore, to use information on reasons why data are missing and incorporate the appropriate dropout mechanism into the data augmentation/analysis.

When missingness is nonignorable, the missingness mechanism must be modeled and inferences should be based on the joint likelihood of the observed data and the missingness mechanism. Wu and Carroll (22) developed a likelihood-based method to handle a class of nonignorable mechanisms termed "informative right censoring." The method involves estimating and comparing rates of changes for missingness due to dropouts. Wu and Bailey (23) derived a computationally simpler version in which the censoring time is used as a covariate in a regression model with individual least squares regression estimates of slopes as the dependent variable. Little's (24) approach assumes that the mean response, over time, of an individual can be modeled as a function of a set of random coefficients, β, and the probability of missingness depends on β. He proposed a class of models called "random coefficient pattern mixture models" to model longitudinal data under the nonignorable "random-effect-dependent dropout" missingness mechanism.

9.6.8.1   Propensity-Adjusted Multiple Imputation Approach

A very useful approach to augmenting informative dropout or truncated data considered here is the *propensity–adjusted multiple imputation* approach (25). This approach utilizes the method of reducing a multivariate stratification to a univariate stratification using the "propensity score" (26, 27). The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. That is, at time t a subject's propensity score is defined as the probability of the subject to remain in the study through time t given the subject's

observed trajectory through time $t-1$. Using the partially observed response vector at time t, $y_t = (y_{t,\text{obs}}, y_{t,\text{mis}})$, covariates x, p_t as the dichotomy indicator for the response y_t ($p_t = 1$ if y_t is completely observed and 0 otherwise), and a logistic regression model, the propensity e_i at time t is given by

$$e_t = (x, y_{0,\text{obs}}, \ldots, y_{t-1,\text{obs}})$$
$$= P(p_t = 1 | x, y_{0,\text{obs}}, \ldots, y_{t-1,\text{obs}}) \quad \text{for } t = 0, \ldots, T \qquad (9.11)$$

As an explanation of the process, assume that data becomes missing at time t—that is, $y_0, \ldots, y_{t-1}$ are completely observed—and that time t is where imputations start. Using the above equation, the propensity scores for time t are calculated and stratified into two propensity strata by splitting them at the median. The observed responses at time t are assigned to these two strata depending on their respective propensity scores. For each propensity stratum, imputations are performed by the approximate Bayesian bootstrap procedure. This consists of selecting responses at random with replacement from the adherent patients to impute for nonadherent patients in the same propensity stratum at time t. The whole process is repeated sequentially with imputed values used as observed values and calculating the propensity of patients based on covariates and their observed trajectories through the previous observations until the end of the study. This process also generates $M \geq 2$ (in practice $M = 10$) possible parallel complete data sets to reflect random variation in the imputation process.

9.6.8.2 Conditional MI

In the validation of a bioanalytical assay for a drug, there is a predefined lower limit of quantification (LLOQ). When determining drug levels in a biofluid, it may be that in later sampling times the concentration of drug in a sample may be below the LLOQ value of the assay. Such a measurement is missing/unknown but it is known that its value is less than the LLOQ and is said to be below the limit of quantification (BQL). Ad hoc approaches have been used to handle such measurements ("left-censored data") during PK analysis. They range from discarding the BQL values to replacing them with either LLOQ/2 or 0. However, a one-value-fits-all approach might lead to infinite weights and problems in the optimization routine, especially the zero substitution. It has been strongly recommended that substituting with 0 should be avoided since it leads to inaccurate and biased parameter estimates (28, 29). We propose a *conditional multiple imputation* (CMI) approach that takes advantage of the properties and advantages of multiple imputation as an alternative for handling BQL data in PK data analysis. With CMI the temporal nature of the response Y, especially in the elimination phase of pharmacokinetics, is maintained. Briefly, the CMI approach involves the assumption of a uniform distribution for BQL observations (i.e., from 0 to LLOQ). A uniform distribution is one for which the probability of occurrence is the same for all values of Y within a range. In particular, the uniform distribution $U(y|a, b)$ on interval (a, b) has mean $(a + b)/2$ and variance $(b - a)^2/12$. In the elimination phase, the range of (a, b) is $(0, \text{LLOQ})$. Assume that data becomes below the lower limit of quantification at time t, that is, $y_t, \ldots, y_n$, are completely unobserved, and time t is where imputations start. The LLOQ observations could be any value between 0 and LLOQ. The

likelihood contribution for such an observation at time k with the expected mean and variance are given below:

$$E(y_k \,|\, y_{k-1}) = y_{k-1}/2, \quad \text{for } k = t, \ldots, n$$
$$V(y_k \,|\, y_{k-1}) = y_{k-1}^2/12 \qquad\qquad (9.12)$$

Thus, the imputation is performed with each imputed observation conditioned on its immediate predecessor. At least 10 replicates of CMIs should be performed for each BQL observation imputed, and the overall PK profile from the replicates should be subjected to analysis and the estimates should be combined as described in Section 9.6.7 to obtain mean parameters and the associated components of variability.

The CMI method assumes that the length of time between the measurement of last quantifiable concentration and the LLOQ value is small relative to the drug half-life. If there is a considerable length of time between the measurement of the last quantifiable concentration and the LLOQ value, then CMI as proposed here has to be modified to account for this. The latter is beyond the scope of this chapter. A motivating example is used to illustrate the application of CMI below.

9.7 A SIMULATION STUDY TO EVALUATE SOME BQL IMPUTATION TECHNIQUES

For the purpose of this investigation, a simulation representing a typical first-time-in-human study was executed. However, only a component of this investigation is reported here. It is typical in early development for the drug assay not to be fully optimized such that the LLOQ is not at the lowest possible value. The consequence is that some subjects may have some (or many) observations that are BQL. In such a setting, data is usually analyzed using the statistical moments approach. The challenge, therefore, is how to estimate noncompartmental PK parameters for the drug under investigation. Thus, the investigation reported here was performed assuming such a setting, and only one dose level is reported here. In addition, only BQL imputation techniques are reported here although several were investigated. The choice is made to provide readers with techniques that have not hitherto been reported.

Median PK parameters for a drug that follows a two-compartment model were used for the simulation (see Table 9.1). Two levels of interindividual variability (30% and 45% coefficient of variation) and two levels of residual variability (15% and 25% coefficient of variation) were evaluated using exponential error models.

TABLE 9.1   Median Pharmacokinetic Parameters Used in the Simulation

PK Parameter	Value
Apparent clearance (CL/F)	29.5 L/h
Apparent volume of distribution of the central compartment (V_2/F)	60.1 L
Apparent volume of distribution of the peripheral compartment (V_3/F)	56.8 L
Intercompartmental clearance (Q)	11.0 L/h
Absorption rate constant (K_a)	3.46 L/h
Absorption lag time (LAG)	0.31 h

The individual parameter values (P_i) were obtained using a lognormal distribution around the population average value (P) according to

$$P_i = P\exp(\eta_i^P) \qquad (9.13)$$

No covariates were included in the model. All simulations were performed in Pharsight Trial Simulator Version V 2.1.2.

Twenty replicates of a typical Phase 1 trial comprising 24 male subjects with an average weight of 70 ± 10 kg and age ranging between 18 and 45 years were simulated for each scenario. Twenty replicates were used in order to determine the number of replicates at which the imputations were stable. The following scenarios were evaluated:

- 30% interindividual variability and 15% residual variability
- 45% interindividual variability and 15% residual variability
- 30% interindividual variability and 25% residual variability
- 45% interindividual variability and 25% residual variability

Hypothetical subjects were generated following administration of a 50 mg dose. The limit of quantification (LOQ) was assumed to be 50 ng/mL. In order to evaluate the performance of the two imputation techniques for imputing BQL concentrations, it is assumed that there are iBLOQ values $[x_1, \ldots, x_i]$. The two methods tested for imputing BLOQ values were CMI and fractional conditional single imputation (FCSI).

CMI values arising from a uniform distribution were used to impute the BQL values as follows:

- The first BQL value (x_1) is imputed assuming a uniform distribution between (0, 50).
- The next BQL value x_2 is then imputed from a uniform distribution between (0, x_1). That is, the imputation of x_2 is conditioned on x_1.
- The process is repeated until all BQL values were generated conditionally.

For FFCSI, BQL values were imputed with LLOQ/n values were $n = 2, 4, 8, 16, \ldots$ LLOQ/n was used instead of LLOQ/2 to ensure that the temporal nature of the profile was maintained.

A true area under the curve (TRUE AUC) was calculated for each subject using the full concentration–time profile generated from the simulation. Sampling was performed at 0, 0.25, 0.5, 0.75, 1, 2, 4, 6, 8, 10, 12, 16, and 24 hours. All areas under the curve were calculated using the noncompartmental analysis module in WinNonlin Version 4.0, using the log/linear trapezoidal rule.

Box plots were generated for each scenario comparing the performance of the two imputation methods. Area under the curve extrapolated to infinity ($AUC_{0\text{-inf}}$), % area extrapolated, and terminal half-life (Lambda Z HL) were plotted and compared across different methods. Also the bias and precision associated with the estimation of each of these parameters were compared for the two methods.

To investigate the stability of the imputation methods with increasing number of replicates, performances of the methods using 5, 10, 15, and 20 replicates were compared. Figure 9.1 shows the performances of the two imputation methods (CMI and FCSI) under the fourth scenario, compared with the full concentration–time profile generated from the simulations. Figure 9.2 shows relative mean prediction errors and the associated standard deviation across 20 replicates for the same scenario. Results for the other scenarios were consistent with Figures 9.1 and 9.2.

Both approaches provide a very good approximation to the true AUC with the bias in the estimation of AUC not exceeding 8% with the CMI method and 10% with the FCSI method (Figure 9.2A). More biased and imprecise estimation of the percent extrapolated AUC was obtained with the CMI method when compared with the FCSI, which yielded a minimally biased and precise estimate of this parameter (Figure 9.2B). The CMI method yielded minimally biased estimate of the terminal elimination half-life while that obtained with the FCSI method was biased. Also, the estimation of terminal elimination half-life was associated with greater imprecision when compared with that obtained with the FCSI method. It should be noted that the FCSI approach is a deterministic approach that does not account
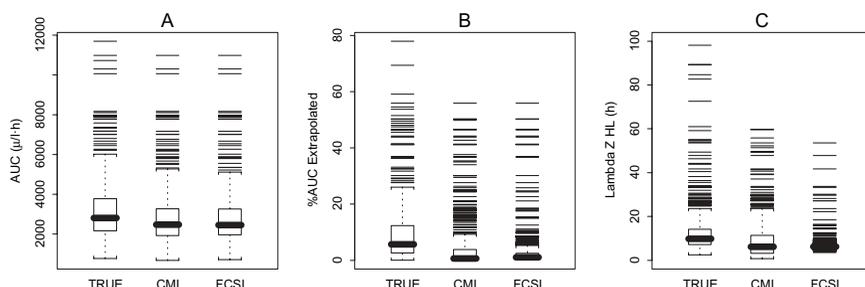

FIGURE 9.1   Performance of the conditional multiple imputation and fractional single multiple imputation (LLOQ/n) methods under the fourth scenario (i.e., assuming 45% interindividual variability and 25% residual variability) for the following parameters: (A) AUC_{0-inf}, (B) %AUC extrapolated, and (C) terminal half-life (Lambda Z HL).
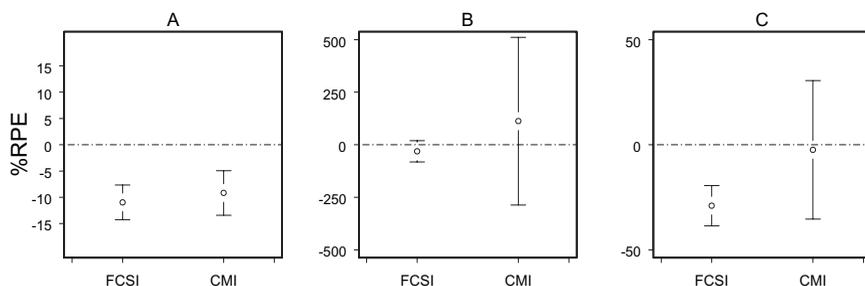

FIGURE 9.2   Bias and precision of the conditional multiple imputation and fractional single multiple imputation (LLOQ/n) methods under the fourth scenario (i.e., assuming 45% interindividual variability and 25% residual variability), presented as percent relative prediction errors (%RPE) (±SD) for the following parameters: (A) AUC_{0-inf}, (B) %AUC extrapolated, and (C) terminal half-life (Lambda Z HL).

for the uncertainty associated with the computation of the terminal elimination half-life (Figure 9.2C). The CMI approach, although more computationally intense, provides a better technique to impute the BQL values when compared with the FCSI approach. However, the FCSI approach could be useful in situations where the uncertainty associated with the estimation of a parameter is not of much primary concern, and this is often the case with noncompartmental analysis. The S-Plus code used to implement the two approaches is shown in Appendix 9.1.

Both imputation methods performed similarly with increasing number of replicates (5, 10, 15, and 20) (results not shown), indicating that these methods are stable with as low as 5 replicates of a particular trial. The similarity in the performance of the two approaches is because the FCSI approach in which imputation was performed with LLOQ/n, where n values were multiples of 2, maintained the temporal nature of the profile. By maintaining the temporal nature of the profile, the FCSI approach in that sense is similar to the CMI approach.

Thus, in a first-time-in-human study where sample sizes may be between 12 and 24, and the objective is to characterize the pharmacokinetics of a new molecular entity, using any of the approaches investigated here would suffice. However, the CMI approach is preferred because it accounts for the uncertainty associated with the imputation of the BQL values while the FCSI approach does not.

9.8   SOFTWARE FOR MI

In addition to the multiple imputation software provided in Appendix 9.1, a general purpose software for performing MI has been written by Schafer (30–33); see also Ref. 8. NORM is a standalone application that performs MI under a multivariate normal model and it is designed for PCs running Windows 95, 98, and NT. The program may be downloaded free of charge from his web site (`http://www.stat.psu.edu/%7Ejls/misoftwa.html#mi`). In addition, four different packages for performing MIs in S-Plus (Insightful Corporation, Seattle, WA) are also available from the same web site. These are NORM (30, 34), which performs MI under a multivariate normal model; CAT (31), for multivariate categorical data under log-linear models; MIX (32), for mixed data sets containing both continuous and categorical data under the general location model; and PAN (33), for multivariate data or clustered data under a multivariate linear mixed effects model.

Harrell (35) developed the transcan function for performing MIs as part of his Hmisc library of miscellaneous S-Plus functions. The transcan function is a general purpose imputation function that can be used for imputing both continuous and categorical variables. The software can be accessed from his web site (`http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/Hmisc?CGISESSID=9ec4d1af802e39c536f7ba4105f5f41b`).

9.9   SUMMARY

Inflation of Type I and Type II error rates, bias parameter estimates, and the degradation of the performance of confidence intervals are possible consequences of improperly addressing the issue of missingness in data analysis. Because a loss of

data is nearly always accompanied by a loss of information, missing values may dramatically reduce statistical power.

In this chapter, different incomplete data types are discussed together with the different imputation techniques that have been developed to impute missing data. Traditionally, incomplete (missing) data have been handled by deletion from analysis of cases that contain missing values (single imputation) and most recently by using MI techniques. The advantages and disadvantages of the different imputation techniques are discussed. MI is the most appropriate approach for imputing missing data because it accounts for the uncertainty associated with the imputation process. Also, the propensity-adjusted MI and CMI are discussed as approaches for handling left-censored data. References for different software available for MI are also provided. Pharmacometricians who wish to mitigate the risks associated with analysis of incomplete data must pay close attention to the issue of missing data in the analysis of clinical trials and choose their strategy carefully.

REFERENCES

1. J. L. Schafer, *Analysis of Incomplete Multivariate Data.* Chapman & Hall/CRC, New York, 2000.
2. D. B. Rubin, Inference and missing data. *Biometrika* **63**:581–582 (1976).
3. E. I. Ette, Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial. *Ann Pharmacother* **32**:818–828 (1998).
4. N. M. Laird, Missing data in longitudinal studies. *Stat Med* **8**:305–315 (1988).
5. R. J. Little and D. B. Rubin, The analysis of social science data with missing values. *Sociol Meth Res* **18**:292–326 (1989).
6. D. B. Rubin, *Multiple Imputation for Nonresponse Surveys.* Wiley, Hoboken, NJ, 1987.
7. P. D. Allison, *Missing Data.* Sage Publications, Thousand Oaks, CA, 2000.
8. J. L. Schafer, Multiple imputation: a primer. *Stat Methods Med Res* **8**:3–15 (1999).
9. B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 4th ed. HarperCollins College Publishers, New York, 2000.
10. L. R. Landerman, K. C. Land, and C. F. Pieper, An empirical evaluation of the predictive mean matching for imputing missing values. *Sociol Meth Res* **26**:3–33 (1997).
11. R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data.* Wiley, Hoboken, NJ, 1987.
12. A. Tsiatis, Analysis and interpretation of trial results: intent-to-treat analysis. *J Acquir Immune Defic Syndr* **3**(Suppl 2):S120–S123 (1990).
13. A. Heyting, J. T. B. M. Tolboom, and J. G. A. Essers, Statistical handling of dropouts in longitudinal clinical trials. *Stat Med* **11**:2043–2061 (1992).
14. D. B. Rubin, Multiple imputation after 18+ years. *J Am Stat Assoc* **91**:473–891 (1996).
15. D. B. Rubin, The Bayesian bootstrap. *Ann Stat* **9**:130–134 (1981).
16. P. D. Allison, *Multiple Imputation for Missing Data: A Cautionary Tale* (1998). Retrieved June 14, 2002, from University of Pennsylvania web site: `http://www.ssc.upenn.edu/ ~allison`.
17. A. S. Bryk and S. W. Raudenbush, *Hierarchical Linear Models.* Sage Publications, Newbury Park, CA, 1992.
18. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London, 1996.

19. M. A. Tanner and W. H. Wong, The calculation of posterior distributions by data augmentation. *J Am Stat Assoc* **82**:528–550 (1987).

20. D. B. Rubin and N. Schenker, Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc* **81**:366–374 (1986).

21. H. Lynn, Maximum likelihood inference for left-censored HIV RNA data. *Stat Med* **20**:33–45 (2001).

22. M. Wu and R. Carroll, Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**:175–188 (1988).

23. M. Wu and K. Bailey, Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**:939–955 (1989).

24. R. J. A. Little, Modeling the drop-out mechanism in repeated measures studies. *J. Am Stat Assoc* **90**:1112–1121 (1995).

25. S. Mazumdar, K. S. Liu, P. R. Houck, and C.F.R. Iii, Intent-to-treat analysis for longitudinal clinical trials: coping with the challenge of missing values. *J Psychiatric Res* **33**:87–95 (1999).

26. P. R. Rosenbaum and D. B. Rubin, The central role of the propensity score in observational studies for causal effect. *Biometrika* **70**:41–45 (1983).

27. P. W. Lavori, R. Dawson, and D. Shera, A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med* **14**:1913–1925 (1995).

28. J. P. Hing, S. G. Woolfrey, D. Greenslade, and P. M. C. Wright, Analysis of toxicokinetic data using NONMEM: impact of quantification limit and replacement strategies for censored data. *J Pharmacokinet Pharmacodyn* **28**:465–479 (2001).

29. S. L. Beal, Ways to fit a PK model with some data below the quantification limit. *J Pharmacokinet Pharmacodyn* **28**:481–504 (2001).

30. J. L. Schafer, *NORM Library for S-PLUS [Computer Software]*. Retrieved Aug. 10, 2005 from `http://www.stat.psu.edu/~jls/misofwa.html` (1997c).

31. J. L. Schafer, *CAT Library for S-PLUS [Computer Software]*. Retrieved Aug. 10, 2005 from `http://www.stat.psu.edu/~jls/misofwa.html` (1997a).

32. J. L. Schafer, *MIX Library for S-PLUS [Computer Software]*. Retrieved Aug. 10, 2005 from `http://www.stat.psu.edu/~jls/misofwa.html` (1997b).

33. J. L. Schafer, *PAN Library for S-PLUS [Computer Software]*. Retrieved Aug. 10, 2005 from `http://www.stat.psu.edu/~jls/misofwa.html` (1997d).

34. J. L. Schafer, *NORM (Version 2.03 for Windows) [Computer Software]*. Retrieved Aug. 10, 2005 from `http://www.stat.psu.edu/~jls/misofwa.html` (1999).

35. F. E. Harrell, *Hmisc Package Website*. Retrieved Aug. 10, 2005 from `http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/Hmisc?CGISESSID=9ec4d1af802e39c536f7ba4105f5f41b`.

APPENDIX 9.1   S-PLUS CODE FOR IMPLEMENTATION OF THE CMI AND FCSI EXAMPLES

```
#Example code to implement methods 1 (CMI) and 2 (FCSI).
# Method 1 (CMI): replace BLOQ values with uniformly distributed
values conditioned to decline based on a preceding value [AUCu]
ct30var$CONCu<-rep(NA,nrow(ct30var))
for (i in 1:20)
{for (j in 1:24)
```

```
   {ct30var.u<-ct30var[ct30var$REP==i & ct30var$PATNO==j,c("REP","P
ATNO","TIME","CONC")]
   ct30var.u$CONCu<-ct30var.u$CONC
   x<-length(ct30var.u$CONC[ct30var.u$CONC<50 & ct30var.u$TIME>2])
   if (x==1)
      {ct30var.u$CONCu[ct30var.u$CONC<50 & ct30var.u$TIME>2]<-
   runif(n=1,0,50)}
      else
      {y<-rep(NA,x)
      y[1]<-runif(n=1,0,50)
        for (k in 2:x)
      {
       y[k]<-runif(n=1,0,y[k-1])
      }
ct30var.u$CONCu[ct30var.u$CONC<50 & ct30var.u$TIME>2]<-y}

if (i*j==1)
{ans30<-ct30var.u}
   else
   {ans30<-rbind(ans30,ct30var.u)}
   cat("\n",i,"REP, and ",j,"PATNO\n")
     }
}
ct30var.u.all<-ans30
###################################################################
#Method 2 (FCSI): replace BLOQ values with
  LOQ/2,LOQ/4,LOQ/8,LOQ/16,LOQ/32,LOQ/64 etc..
ct30var$CONCu<-NULL
ct30var1<-ct30var
ct30var1$CONC1<-ct30var1$CONC

for (i in 1:20)
{for (j in 1:24)

  {
    ct30var.1<-ct30var1[ct30var1$REP==i &
    ct30var1$PATNO==j,c("REP","PATNO","TIME","CONC", "CONC1")]

    ct30var.1$CONC1[ct30var.1$CONC<50 & ct30var.1$TIME>2]<-
    c(25,12.5,6.25,3.125,1.56,0.78)

  if (i*j==1)
    {ans1<-ct30var.1}
    else
    {ans1<-rbind(ans1,ct30var.1)}
cat ("\n",i,"REP, and ",j,"PATNO\n")
  }
}ct30var.u1.all<-ans1
```

POPULATION PHARMACOKINETIC BASIS OF PHARMACOMETRICS

Population Pharmacokinetic Estimation Methods

ENE I. ETTE, PAUL J. WILLIAMS, and ALAA AHMAD

10.1   INTRODUCTION

Population pharmacokinetics is the study of the sources and correlates of variability in drug concentrations among individuals who are the target patient population receiving clinically relevant doses of a drug of interest (1). Certain patient demographical, pathophysiological, and therapeutic features, such as body weight, excretory and metabolic functions, and the presence of other therapies, can regularly alter dose–concentration relationships. Population pharmacokinetic (PPK) models are vital to the drug development and evaluation processes by providing predictions of the individualized dose–exposure relationship, which is pivotal to rational and successful drug therapy. There are several advantages to employing PPK models when compared to traditional pharmacokinetic (PK) model development (2). Unlike the traditional studies in which subjects are sampled intensively, the population approach to studying the pharmacokinetics of a drug allows both sparsely and intensively sampled data to be used. It enables the execution of PK studies in special populations such as neonates (3, 4), elderly (5, 6), AIDS patients (7), critical care patients, and cancer patients (3), where the number of samples to be obtained per subject are limited because of ethical and medical concerns. During drug development relatively few samples can be obtained from subjects/patients participating in Phase 2 and 3 studies for the determination of the pharmacokinetics of a drug in the relevant population and for the determination of the relationship between dose, exposure (concentration), and response/safety.

The sparse sampling approach for characterizing population pharmacokinetics yields better estimates of intersubject variability than traditional approaches (discussed later), which yield positively biased estimates of this measure of dispersion (8–10). A combination of accurate and precise estimates of intersubject variability and the mean parameter value for a drug is useful for selecting an initial dose strategy for drug therapy in a patient and allows Bayesian feedback analysis to be performed for dosage individualization.

The analyses of sparse samples collected for PPK analysis have been reported to be cost effective (11). The analytical cost for samples for Phase 2 and 3 trials is reported to be typically comparable to the total cost of a single Phase 1 special topic study (11). If an average of 3 samples are taken from 1000 patients and the analytical cost is $80 (US) per sample, the $240,000 (US) analytical expenditure will typically allow quantitative assessment of multiple factors affecting the drug's pharmacokinetics, including such factors as gender, race, age, renal function, concomitant drug use, and disease severity. Database support, overheads, and pharmacokineticist's time probably account for $60,000–100,000 (US), depending on the complexity of the analyses. The savings are obvious when this is compared to an approximate $450,000 for a 10 day multiple-dose crossover drug–drug interaction study in 18 healthy volunteers.

PPK analyses provide an opportunity not only to estimate variability but also to identify its sources. Variability is usually characterized in terms of fixed and random effects. The fixed effects are the population average values of PK parameters, which may in turn be a function of patient characteristics discussed earlier. The random effects quantify the amount of PK variability, which is not explained by the fixed effects, and these random effects subsets are intersubject variability, interoccasion variability, and intraindividual and residual variability (12). Karlsson and Sheiner (13) have demonstrated that the lack of separation of interoccasion variability from residual intrasubject variability produced bias in estimation of PPK parameters.

The PPK approach can allow one to combine heterogeneous types of data from varying sources. For example, one could pool data from several different studies, study centers, variable biomatrices (plasma plus serum), intensely plus sparsely sampled data, or experimental plus observational data. The combining of differing data sets often increases the power to identify multicompartment or nonlinear models, to incorporate additional covariates, or to gain precision in the estimation of the model.

A disadvantage of the PPK approach is that it requires skilled pharmacokineticists/pharmacometricians who are able to implement the mathematical and statistical techniques used in the estimation of PPK parameters.

Over the past 25 years a variety of methods have been proposed for the characterization of the population pharmacokinetics of drugs. In this chapter, the statistical framework for estimating population pharmacokinetics in terms of individual and population models is discussed as a prelude to discussing some of the methods used in estimating population pharmacokinetics. In doing so we have adopted a user-friendly approach described previously (14). The goals of a PPK analysis and the data type (1) will determine the method selected for the analysis.

10.2   STATISTICAL FRAMEWORK FOR ESTIMATING POPULATION PHARMACOKINETICS

10.2.1   Individual Model

When engaging in PPK model development, it is important to understand the underlying principles of both the individual and the population models. This avoids working with the black box mentality, which can lead to the estimation of inappropriate models.

The PK model for estimating individual PK parameters can be written

$$y_j^m = f_j(\phi_j, x_j) \tag{10.1}$$

where $y^m{}_j$ is the model predicted vector of values of y_j (an observed dependent variable) when the PK parameters take the value ϕ_j, and where f_j stands for the functional relationship (i.e., PK model) between the predictions and the ϕ_j. x_j is also a vector of known quantities (dose D, time t_j, subject size, etc.).

Assuming measurement errors to be zero, Eq. (10.1) would be rewritten as follows:

$$y_j = f_j(\phi_j, x_j) \tag{10.2}$$

if the model-predicted values, $y^m{}_j$, are equal to the measured values y_j. In practice, however, the observations must be related to individual true parameters ϕ_j through a somewhat more elaborate model:

$$y_j = f_j(\phi_j, x_j) + \varepsilon_j \tag{10.3}$$

where the vector ε_j is a sample of the measurement noise E_j.

Maximum likelihood (ML) estimation can be performed if the statistics of the measurement noise E_j are known. This estimate is the value of the parameters for which the observation of the vector, y_j, is the most probable. If we assume the probability density function (pdf) of E_j to be normal, with zero mean and uniform variance, ML estimation reduces to ordinary least squares estimation. An estimate, $\phi^*{}_j$, of the true jth individual parameters ϕ_j can be obtained through minimization of some objective function, $O_j(\theta_j)$. The model given by Eq. (10.3) is assumed to be a natural choice if each measurement is assumed to be equally precise for all values of y_j. This is usually the case in concentration–effect modeling.

There are instances where the error changes with differing values of y_j. When this is the case, one ought to consider the multiplicative lognormal error model, where the observed concentration y is given by

$$y_j = f_j(\phi_j, x_j)e^{\varepsilon_j} \tag{10.4}$$

where ε_j is assumed to follow a lognormal distribution with median 1 and constant coefficient of variation CV. The variance of the prediction of the deviations is proportional to the square of the predicted response. The lognormal error model is often appropriate if measurements can only be positive and if they become less precise when the measured value increases. This is often true of biological systems in general and PK models in particular. However, other error models may be used in practice such as a combination of the additive and lognormal error model to improve prediction at the lower limit of assay precision, where variance may be assumed constant, and the power function model where the variance of the deviations is assumed proportional to some power (which is to be estimated) of the predicted response.

When the estimation procedure is clearly specified, an approximate covariance matrix of the estimate, S_j, can also be calculated. This matrix reflects the degree

of precision of the estimate and depends on the experimental design, parameters, and noise statistics. A well designed experiment with small random fluctuations will lead to precise estimations ("small" covariance), while a small number of uninformative data and/or a high level of noise will produce unreliable estimates ("large" covariance).

10.2.2 Population Model

Unlike the individual model discussed above, a more elaborate statistical model is required to deal with sparse PK data. In formulating the model, it is recognized that overall variability in the measured (response) data in a sample of individuals reflects both measurement error and intersubject variability. The observed response (e.g., concentration) in an individual within the framework of population (regression) nonlinear random mixed effects models can be described as

$$y_{ij} = f_{ij}(\phi_{ij}, x_{ij}) + \varepsilon_{ij} \tag{10.5}$$

where y_{ij} for $i = 1, \ldots, n_j$ are the observed data (e.g., blood or urinary levels at time points x_{ij}) of the jth subject. The model given by Eq. (10.5) is defined for all $j = 1, \ldots, N$, where N is the number of subjects in the sample. f_{ij} is a specified function for predicting the ith response in the jth subject (e.g., one or several exponentials), ε_{ij} is the ith measurement error in the jth subject, and ϕ_j was previously defined (15). From the notations, it is clear that the population model is a collection of models for individual observations. Various doses and/or dosage regimens and/or administration routes are generally used in patient drug therapy and clinical trials. Also, diverse administration schedules (single dose and multiple dosing) might be used and several responses (e.g., plasma and urinary drug levels) might be measured. Correspondingly, in drug-related "population" applications, the functions f_{ij} will differ across individuals. In contrast, it is realistic to assume that the set of underlying structural parameters, in this case PK parameters, is qualitatively the same for all individuals and that the parameters vary quantitatively among individuals. Mathematically this can be written

$$\phi_j = g(\theta, z_j) + \eta_j \tag{10.6}$$

where g is a known function that describes the expected value of ϕ_j as a function of known individual specific covariates z_j, such as weight, age, disease state, and concomitant medication, and the vector of population parameters θ. Covariates are assumed constant within an individual, for simplicity. However, time-varying covariates can also be incorporated in the model by permitting individual PK parameters to depend on i as well as on j. η_j represents the random variation of the individual parameter vectors around the population prediction. The η_j are usually assumed to be independent across individuals (i.e., η_j, η_l are independent for $j \neq l$). The individual ϕ_j is assumed to arise from some multivariate probability $F(\theta)$.

In the mixed effects context, the collection of population parameters is composed of a "population-typical value," generally the mean, and of a "population-variability value," generally the variance–covariance matrix. The mean and variance are the first two "moments" of a probability distribution. They build a minimal set of hyper-

parameters or "population characteristics" for it, which is sufficient in a statistical sense when F is taken as normal or lognormal.

Thus, a PK model that describes the time course of the drug in the body in a specific individual, a model describing the relationship between patient characteristics and the PK model parameters, a variance model for residual random variability, and a population model for intersubject random variability that describes the unexplained random variability of the model parameters in the population of subjects studied are essential for describing a PPK model.

10.3 METHODS APPLIED TO POPULATION PHARMACOKINETIC MODELING

10.3.1 Naive Average Data Approach

It is common practice in preclinical and clinical pharmacokinetics to perform studies in which the drug administration as well as the sampling schedules are identical for all subjects. For this type of analysis there are as many data points as there are individuals at each sampling time. Analysis of such data using the naive averaging of data (NAD) approach consists of the following procedure:

1. Compute the average value of the data for each sampling time

$$\bar{y}_i = \frac{1}{N}\sum_{j=1}^{N} y_{ij}$$

(10.7)

   for $i = 1, \ldots, n$, where n is the standard number of individual data. The averaging of data across individuals makes sense, because all y_{ij} for $j = 1, \ldots, N$ have been measured under identical conditions.
2. A model $y^m = f(\phi)$ is fitted to the mean-data n-vector $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_n)^t$ while estimating the best-fit parameter values ϕ^*. The latter notation (ϕ^*) is used to distinguish it from individual estimates, denoted $\hat{\phi}$.

The NAD approach is attractive because of its simplicity. One unique fitting is sufficient for obtaining estimates of parameters describing the mean response. ϕ^* components are quite often interpreted as "mean" parameter values. Correspondingly, $\hat{\mu}_{NAD}$ will be used for ϕ^* in the latter. The method is widely applicable in experimental data (EP) studies with standardized designs, and examples of these include bioavailability, bioequivalence, and dose proportionality studies. Because of the smoothing effect of averaging, mean data generally look nicer than individual data, and better fitting often results when compared with individual data.

However, the NAD approach provides an estimate of $\hat{\mu}_{NAD}$ sample mean. In this regard, several drawbacks of this approach must be pointed out. The use of NAD to establish a PK model may be misleading. Quite often, data averaging can produce a distorted picture of the response. Averaging of monoexponential data from two subjects with very different half-lives has been shown to produce a mean curve that exhibits an apparent biexponential decay (16). Sometimes the opposite situation is

the case. The smoothing effect of the averaging will tend to obscure peculiarities that can be seen in individual data. The existence of secondary peaks in the plasma level–time course of individuals may be undetectable in the average curve if the rebounds occur at different time points.

NAD also performs poorly in terms of parameter estimation. The reference to individual data disappears after data averaging. All sources of variability are confounded. Because of this, important information on drug disposition is obscured. The average concentration curve derived with the NAD approach does not necessarily follow the individual model function. A wrong model may be obtained (17). Undefined statistical uncertainties and large "unknown" subject variations might smooth the average response curve in an unpredictable manner. Thus, the NAD estimate $\hat{\mu}_{\mathrm{NAD}}$ should not, as a general rule, be regarded as a valuable estimate of the expected value of PK parameters. This rule holds even if the true model, that is, the one that adequately describes the individual data, has been used for the fitting. The essential parametric nonlinearity of PK models is responsible for this. Exceptions to this rule occur when the signal-to-noise ratio is small. This is the case when variability contributes less to the spread in observations than other sources of fluctuation (interoccasion variability, measurement error, and model misspecification). This situation might be seen when concentrations are measured in standardized laboratory animals. The quality of estimates may be improved by using averaging methods other than straightforward arithmetic mean (18). These ad hoc solutions do not fundamentally solve the problem. Moreover, no estimate of pure interindividual variability can be obtained with the NAD approach because it masks variability rather than revealing it. Thus, the NAD approach is not a reliable method for PK data analysis.

10.3.2 Naive Pooled Data Analysis

Sheiner and Beal (19) proposed the term naive pooled data (NPD) approach for the method in which all data from all individuals are considered as arising from one unique individual. This reference subject is characterized by a set of parameters ϕ. With least-squares fitting, ϕ will be the parameter vector minimizing the global objective function

$$O_{\mathrm{NPD}}(\phi) = \sum_{j=1}^{N} \left\{ \sum_{j=1}^{nj} [y_{ij} - f_{ij}(\phi)]^2 \right\} \tag{10.8}$$

where $\{f_{ij}, i = 1, \ldots, n_j\}$ is the set of components of f_j, and the summation is over all individuals and all measurements for a given individual.

Unlike the NAD approach, the NPD approach is far more general. It can easily deal with experimental data, nonstandard data, and routine PK data. After a unique fitting of all data at once, parameter estimates are obtainable. It may perform well when variations between subjects are small. This is occasionally the case in a group of homogeneous laboratory animals from a given strain, but it is rarely true for humans. The drawbacks of NPD are the same as those of NAD, as has been repeatedly pointed out (20–22). The NPD approach tends to confound individual differences and diverse sources of variability in a manner different from the NAD

approach, but with similar negative consequences. The NPD estimate for the reference individual ϕ should be considered as a rough approximation ($\hat{\mu}_{NPD}$) of the population expectation μ, although the consequences of the omission can be minor (23). In addition, estimates of the dispersion of parameters in the population are not provided. Extrapolation of mean outcomes on the basis of the set of estimates $\hat{\mu}_{NPD}$ should be done with caution.

These problems notwithstanding, it has been shown by Shafer and co-workers (24–26) that for several drugs used in anesthesia a pooled analysis approach provided population mean parameters that, when prospectively tested, accurately predicted drug concentrations after drug administration by a computer-controlled infusion pump. The data, in all circumstances, originated from well controlled experiments with extensive sampling. That is, the data were of the experimental data (EP) type. Moreover, the NPD analysis provided similar population mean parameter estimates when compared with estimates obtained using several other population analysis methods (27, 28). These findings are in contrast with an earlier simulation study that showed that the NPD approach provided biased estimates of the population mean parameters even when a well balanced experimental study design was used (20). The discrepancy may be due to the large amount of interindividual variability present or inappropriate weighting scheme used in the latter study.

Imbalance and confounding correlations present in a data set pose serious problems for the NPD approach. These features are prevalent in observational data and make the NPD approach inappropriate for this type of data. Data imbalance occurs when there are many more observations taken from some individuals than others. An example would be a case where six samples are taken from individuals, four from some, and one from others.

When the design of the study correlates with the outcome, confounding correlations occur. That is, the presence or absence of an observation is dependent on the subject's pharmacokinetics. Confounding correlations are usually prevented with randomization. This, however, is not guaranteed with observational data. A case in point would be a PK study in which concentrations fall below the limit of quantification during the study. Only individuals with the smallest clearance or largest volume of distribution would contribute measurable concentrations toward the end of the study. Biased estimate of the terminal half-life will result and may be wrongly interpreted as an additional phase of the PK profile. Clearly, the NPD approach should not be used in this setting.

10.3.3 Two-Stage Approach

With this approach, individual parameters are estimated in the first stage by separately fitting each subject's data, then in the second stage obtaining parameters across individuals, thus obtaining population parameter estimates. The data are summarized in the set $[(\hat{\phi}_j, M_j), j = 1, \ldots, N]$. $\hat{\phi}_j$ is the p-vector of the parameter estimates and is the $p \times p$ symmetric variance–covariance matrix of the corresponding individual estimate. To derive values for population characteristics according to a given strategy, the individual parameter estimates are combined. The salient features of the methods that constitute the two-stage approach are discussed briefly.

10.3.3.1   *Standard Two-Stage (STS) Approach*

The STS approach refers to a well known and widely used procedure. Population characteristics of each parameter are estimated as the empirical mean (arithmetic or geometric) and variance of the individual estimates $\hat{\phi}_j$ according to the following equations:

$$\hat{\mu}_{\text{STS}} = \frac{1}{N}\sum_{j=1}^{N}\hat{\phi}_j \tag{10.9}$$

$$\hat{\Omega}_{\text{STS}} = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\phi}_j - \hat{\mu}_{\text{STS}}\right)^2 \tag{10.10}$$

The estimate of the standard deviation ($\hat{s}$) is easily obtained by taking the square root of $\hat{\Omega}$. $N - p$ can be used instead of N in the denominator of the variance estimate.

With the STS approach estimates of individual parameters are combined as if the set of estimates were a true N-sample from a multivariate distribution. It has been recommended as a very simple and valuable approach for pooling individual estimates of PK parameters derived from experimental PK studies (29). The advantage of the STS approach is its simplicity, but the validity of its results should not be overemphasized. However, it has been shown from simulation studies that the STS approach tends to overestimate parameter dispersion (the variance–covariance matrix) (20, 30).

10.3.3.2   *Global Two-Stage (GTS) Method*

The $\hat{\phi}$ can be viewed as observations of the individual parameters. The estimate for a subject may be biased and imprecise because of poor experimental design, poor study execution, or a high level of measurement error. The GTS approach makes extensive use of the matrices $|M_j, j = 1, \ldots, N|$, which reflect the deviations (bias), together with the estimates $|\hat{\phi}_j, j = 1, \ldots, N|$. The expectation $E(\cdot)$ and the variance–covariance $Var(\cdot)$ of each (random) $\hat{\phi}_j$ can be calculated:

$$E\left(\hat{\phi}_j\right) = \mu \quad \text{for} \quad j = 1, \ldots, N \tag{10.11}$$

$$\text{Var}\left(\hat{\phi}_j\right) = M_j + \Omega \quad \text{for} \quad j = 1, \ldots, N \tag{10.12}$$

where μ is the true population expectation and Ω is the true population variance–covariance. An extensive description of the method is provided by Steimer et al. (30). The GTS approach provides a maximum likelihood estimate of μ and Ω by an iterative method. It assumes that the estimates of individual parameters are normally distributed around the true parameters with variance Var_j. The population parameters θ are the p components of the vector μ and the $p(p + 1)/2$ independent components of the symmetric matrix Ω. The objective function to be minimized is as follows:

$$O_{\text{GTS}}(\mu, \Omega) = \sum_{j=1}^{N}\left[\left(\hat{\phi}_j - \mu\right)^t (M_j + \Omega)^{-1}\left(\hat{\phi}_j - \mu\right) + \ln\det(M_j + \Omega)\right] \tag{10.13}$$

The first term on the right-hand side of Eq. (10.13) is the summation (over individuals) of the weighted squared deviations of individual estimates from the expected value μ. The weighting matrix is dependent on the quality of the estimate through the factor $(M_j + \Omega)^{-1}$. The last term in the equation is the logarithm of the determinant of the $(M_j + \Omega)$ matrix. It prevents the variance–covariance matrix from going to zero through its determinant.

The GTS approach has been shown, through simulation, to provide unbiased estimates of the population mean parameters and their variance–covariances, whereas the estimates of the variances were upwardly biased if the STS approach was used (30). These simulations were done under the ideal situation that the residual error was normally distributed with a known variance. However, it is a well known fact that the asymptotic covariance matrix used in the calculations is approximate and under less ideal conditions that the approximation can be poor (31, 32).

10.3.3.3 *Iterative Two-Stage (IT2S) Approach*

A computationally "heavier" two-stage method that relies on repeated fittings of individual data (IT2S) has been described (30, 33, 34). The IT2S approach can be implemented with rich data, sparse data, or a mixture of both. An approximate a priori population model is required to initiate the procedure. Provided that considerable informative data is available, the population values may be obtained from the literature, the NPD approach can be performed with the current study data with a reasonable choice of parameter variability, or the STS approach (30). As the name implies, the IT2S approach is implemented in two stages. In the first stage, the population model is used as the set of prior distributions for Bayesian estimation of the individual parameters for all patients, irrespective of the number of samples supplied by each individual. In the second stage, the population parameters are recalculated with these new individual parameters in order to form the new set of prior distributions. The estimation process (i.e., parameters from the second stage are used for a repeat of the first stage and the results are used for a repeat of the second stage) is repeated until the difference between the new and old prior distributions are essentially zero. The method may be implemented with programs supporting Bayesian estimation and least-squares regression or with the IT2S routine (34), which has been implemented with the USC*PACK collection of programs (35).

A method close to the IT2S procedure is the expectation-maximization-like (EM) method presented by Mentre and Geomeni (36), which can be viewed as an extension of the IT2S procedure when both random and fixed effects are included in the model and for heteroscedastic errors known to a proportionality coefficient. This algorithm is implemented with the software P-PHARM (37).

10.3.3.4 *Bayesian Two-Stage Approach*

A method that is Bayesian in nature is that proposed by Racine-Poon (38). The method uses the estimates of the individual parameters ϕ_j and asymptotic variance matrix V_j obtained from the individual fits, with very weak assumptions about the prior distribution of the population parameters to calculate a posterior density function from which ϕ and Ω can be obtained. In an iterative method suggested by Dempster et al. (39) the EM algorithm is used to calculate the posterior density function. Simulation studies in which several varying and realistic conditions were

assumed have shown that the Bayesian two-stage approach provides good estimates of PPK and pharmacodynamic (PD) parameters (38, 40).

10.3.3.5 Bayesian Analysis Using Gibbs Sampling

BUGS (*B*ayesian inference *u*sing *g*ibbs *s*ampling) is a general program involving a fully Bayesian approach (41). The program is an implementation of the Markov chain Monte Carlo (MCMC) method (42, 43). An increasingly used Bayesian software for PPK data analysis is WinBUGS, which includes an add-on for PK analysis (PKBUGS). Chapter 5 covers Bayesian methods and the use of WinBUGS.

10.3.4 Nonlinear Mixed Effects Model Approach

The first attempt at estimating interindividual PK variability without neglecting the difficulties (data imbalance, sparse data, subject-specific dosing history, etc.) associated with data from patients undergoing drug therapy was made by Sheiner and co-workers (44) using the nonlinear mixed-effects model approach. The vector θ of population characteristics is composed of all quantities of the first two moments of the distribution of the parameters: the mean values (fixed effects) and the elements of the variance–covariance matrix that characterize random effects (19, 20, 45–47).

The number of samples per subject used for this approach is typically small, ranging from one to six. The difficulties associated with this type of data preclude the use of the STS approach because there are not enough data to estimate the PK parameters for each subject separately. There are too few measurements to estimate the parameters accurately or the model may be unidentifiable in a specific individual. As does the pooled analysis technique, nonlinear mixed effects modeling approaches analyze the data of all individuals at once but take the interindividual random effects structure into account. This ensures that confounding correlations and imbalance that may occur in observational data are properly accounted for.

Most of the nonlinear mixed effects modeling methods estimate the parameters by the maximum likelihood approach. The probability of the data under the model is written as a function of the model parameters, and parameter estimates are chosen to maximize this probability. This amounts to asserting that the best parameter estimates are those that render the observed data more probable than they would be under any other set of parameters.

It is difficult to calculate the likelihood of the data for most PK models because of the nonlinear dependence of the observations on the random parameters η_i and possibly ε_{ij}. To deal with these problems, several approximate methods have been proposed. These methods, apart from the approximation, differ widely in their representation of the probability distribution of interindividual random effects.

10.3.4.1 First-Order Method

The first nonlinear mixed effects modeling program introduced for the analysis of large amounts of PK data was NONMEM (48). In the NONMEM program linearization of the model in the random effects is effected by using the first-order Taylor series expansion with respect to the random effect variables η_i and ε_{ij}. This software is the only program in which this type of linearization is used.

The jth measurement in the ith subject of the population can be obtained from a variant of Eq. (10.5) as follows:

$$y_{ij} = f(\phi, x_{ij}, \eta_i) + \varepsilon_{ij} \tag{10.14}$$

The first-order Taylor series expansion of the above model with respect to the random variables η_i (intersubject variability) and ε_{ij} (residual variability) around zero is given by

$$y_{ij} = f(\phi, x_{ij}) + G_{ij}(\phi, x_{ij})\eta_i + \varepsilon_{ij} \tag{10.15}$$

where

$$G_{ij}(\phi, x_{ij}) = \delta f(\theta, x_{ij}\ \eta_i, \varepsilon_{ij})/\delta\eta_i^T \,|_{\eta_i=0} \tag{10.16}$$

$G_{ij}(\phi, x_{ij})$ is a $1 \times p$ matrix of the first derivatives of $f(\theta, x_{ij}\eta_i, \varepsilon_{ij})$ with respect to η_i, evaluated at $\eta_i = 0$. In Eq. (10.15) the model is linear in ε_{ij}; therefore, no approximation is made with respect to ε_{ij}. Logarithmic transformation of the data can be performed to ensure linearity in ε_{ij}.

The random effect parameters η_i and ε_{ij} are independent (multivariate) normally distributed with zero means and variances Ω and σ^2, respectively. Ω is the $p \times p$ covariance matrix of the p vector η_i. Based on the fact that η_i and ε_{ij} are independent identically normally distributed, and the linearization of Eq. (10.15), the expectation and variance–covariance of all observations for the ith individual (first two moments) are given by

$$E_i = f(\theta, x_i) \tag{10.17}$$

and

$$C_i = G_i\big((\theta, x_i)\Omega G_i(\theta, x_i)^T\big) + \sigma^2 I_{ni} \tag{10.18}$$

where $f(\theta, x_i)$ is the vector of model predictions of y_i, $G_i(\theta, x_i)$ represents the $n_i \times p$ matrix of first derivatives of $f(\theta, x_i, \eta_i, \varepsilon_i)$ with respect to η_i evaluated at $\eta_i = 0$, and I_{ni} represents the identity matrix of size n_i. Maximum estimates of the population parameters θ, Ω, and σ^2 can be obtained by minimizing minus twice the logarithm of population likelihood as expressed below:

$$-2LL = \sum_{i=1}^{N} \big(\log(\det(C_i)) + (y_i - E_i)^T C_i^{-1}(y_i - E_i)\big) \tag{10.19}$$

This approach is called the first-order (FO) method in NONMEM. This is the most widely used approach in PPK and PD data analysis and has been evaluated by simulation. The use of the first-order Taylor series expansion to approximate the nonlinear model in η_i and possibly ε_{ij} by a linear model in these parameters is the greatest limitation of the FO approach.

The performance of the FO approach for the analysis of observational and experimental data has been evaluated by Sheiner and Beal (19) with the

Michaelis–Menten PK model and the one- and two-compartment models (20, 21). In all instances, a comparison was made with the naive pooled data and standard two-stage approaches for the analysis of the two types of data. The FO approach outperformed the NPD and the STS approaches on both data types. Despite the approximation, the FO approach provides good parameter estimates. When the residual error increases, the STS approach quickly deteriorates, especially with respect to variance parameters. However, the STS approach still performs reasonably well but the bias and imprecision of the estimates tend to increase with increasing residual error (21). Estimates of residual random effects have been shown to deteriorate with the FO approach when residual error increases (49).

Deterioration in parameter estimation has been observed in simulation studies in which the value of the intersubject variability was greater than 60% and the residual variability was set at 15% (50). A series of studies in which observations were randomly deleted from a data-rich set to create a sparse data set, and parameter estimation done using the FO showed good performance of the FO approach when compared with the results obtained using the full data set (51–55). The correspondence of the results in the two situations suggests that the FO approach can be used to estimate parameters using only a few observations per individual. Simulation studies have been performed to show that the FO approach can be used in the limiting case where only one sample is obtained per subject (56). In this case, there is an upper limit of residual variability (not exceeding 20%) for the production of reliable parameter estimates.

The impact of the linearization approximation of the FO approach for a simple one-compartment model was evaluated by Beal (46). He compared the performance of this approach with the exact solution to the population likelihood. No difference was observed, which indicated that the approximation used in the FO method is not detrimental to the analyses under the conditions evaluated, which included an interindividual variability set at 25% (CV%). Other simulation studies, however, have shown that the FO approach has a potential for providing modestly biased estimates (20, 33, 45, 51, 57–60).

For a one-compartment multidose scenario, White et al. (32) showed that biased estimates are more likely when residual and intersubject variabilities are very high. Ette et al. (50) observed that the biased estimates are obtained at high levels of intersubject variability with a two-compartment multidose situation although the residual variability did not exceed 15%. The bias may be due to the fact that the first-order Taylor series expansion is not a particularly good approximation of the underlying "real" (lognormal) distribution used to generate the simulated data in these studies. Also, it may be that the first-order Taylor series expansion is evaluated at $\eta_i = 0$ (the population mean estimate of η_i). This may not be a good approximation depending on the magnitude of intersubject variability and the nonlinearity of the pharmacokinetic model. During data analysis, this can be compensated for, in part, by including explanatory covariates in the model to reduce the variance of η_i. With a one-compartment model experimental data set, the GTS approach was shown to outperform the FO approach with respect to bias and precision of both the population mean and variance estimates. Similar results were obtained in a study in which the FO approach was compared with the Bayesian two-stage approach (57).

The NONMEM program implements two alternative estimation methods, the first-order conditional estimation (FOCE) and the Laplacian methods (48). The FOCE method uses a first-order expansion about conditional estimates (empirical Bayes estimates) of the interindividual random effects, rather than about zero (61). In this respect, it is like the conditional first-order method of Lindstrom and Bates (62). Unlike the latter, which is iterative, a single objective function is minimized, achieving a similar effect as with iteration. The Laplacian method uses second-order expansions about the conditional estimates of the random effects (61).

10.3.4.2   Conditional First-Order (NLME)
The conditional first-order method of Lindstrom and Bates uses a first-order Taylor series expansion about conditional estimates of interindividual random effects (62). Estimation involves an iterative generalized least-squares type algorithm. This estimation method is available in S-Plus (Insightful, Seattle, WA) as the function NLME (63).

10.3.4.3   Alternative First-Order (MIXNLIN)
This method, proposed by Vonesh and Carter (64), also uses a first-order series expansion of the interindividual random effects. They proposed the use of estimated generalized least squares and established the asymptotic properties of the resulting estimates. An alternative method is the use of the iteratively reweighted generalized least squares (65). The MIXNLIN program also implements pseudo maximum likelihood (ML) and restricted maximum likelihood (REML) estimation by embedding the EM algorithm within an iteratively reweighted generalized least-squares routine. Expansion is either about zero or about the empirical best linear unbiased predictor (EBLUP) of the interindividual random effects. Only the fixed effects and variance component estimates are updated after each call to the embedded EM algorithm (i.e., the method uses the EBLUP estimates inherent within the EM algorithm only to update estimates of the variance components) when the expansion is about zero. Maximum likelihood estimation expanded about zero should result in estimates similar to those obtained using the NONMEM first-order method, while expansion about the EBLUP should result in estimates similar to those obtained with the first-order conditional estimation in NONMEM and by the first-order conditional method (NLME). These estimation methods are available in the SAS macro and MIXNLIN 3.0 of Vonesh (65).

10.3.4.4   Alternative First-Order (SAS)
This is a first-order Taylor series expansion method, but the algorithm consists of iteratively fitting a set of generalized estimating equations until they stabilize (66). The method uses a Taylor series expansion in the fixed effects parameters, as well as one in the random effects; expansion is about the generalized least-squares estimates for the fixed effects parameters, and about zero for the random effects. It yields estimates similar to those obtained using the first-order method of NONMEM. The method is implemented in the SAS macro NLINMIX. The NLINMIX program also implements expansion about the EBLUPs of the interindividual random effects, as an alternative to expansion about zero, yielding estimates similar to those produced with the FOCE method in NONMEM.

10.3.5   Nonparametric Maximum Likelihood (NPML)

The NPML approach provides an estimate of the whole probability distribution of the PK parameters on a nonparametric basis (67). The method relies on maximization of the likelihood of the set of observations of all individuals to estimate the distribution of the parameters. The basic conceptual framework is similar to that described for NONMEM above. The difference is that no specific model for the relationship between PK parameters and patient-specific covariates is specified. The individual parameters ϕ_i are assumed to be independent realizations of a given random variable Φ with probability distribution $F(\phi)$. The likelihood of all data is given by

$$L(F) = \prod_{i=1}^{N} \int {}_D l_i(y_i | \phi) F(\phi) d\phi \qquad (10.20)$$

where $l_i(y_i|\phi)$ is the likelihood of the observations y_i for ith individual, given ϕ. D is the domain in which the parameters lie. Maximization of this likelihood provides an estimate $\hat{F}$ of the probability distribution of the parameters. This distribution has been proved by Mallet (67) to be discrete, involving N_p locations, where N_p is less than or equal to the number of individuals (N). To estimate the N_p locations q_k and their corresponding frequencies α_k, a specific algorithm was developed. The level of residual error and how well the parameters are known determines the number of locations. There will be N locations, each with a frequency of $1/N$ if the parameters are known very precisely for all N subjects. The set of locations q_k and frequencies α_k completely specifies the estimate of the distribution of the parameters:

$$\hat{F} = \sum_{k=1}^{N_p} \alpha_k \delta(q_k) \qquad (10.21)$$

where $\delta(x)$ denotes the Dirac probability distribution, which takes the value 1 at x and 0 elsewhere. With this method a complete distribution of F with very soft assumptions, namely, that F takes only positive values and that its integral over domain D is equal to unity (67, 68). The NPML approach has been shown in a simulation study, assuming a one-compartment PK model with bimodal distribution, to produce parameter estimates that accurately describe the distribution, even though only one measurement was available per individual (69). Several summary statistics such as mean or variance–covariance matrix can easily be calculated from the distribution of F specified by Eq. (10.21). The method also allows for the inclusion of patient-specific covariates without specifying the a priori relationship between the PK parameters and covariates. The covariates are regarded as additional parameters and the algorithm provides an estimate of the joint distribution of the PK parameters and the covariates (70). The probability distribution of the parameters conditional on any value of the covariates can be computed and used for the initial dosage selection, given the distribution obtained. Thus, the shape of the relationship between parameters and covariates can be explored nonparametrically.

   The major limitation of this approach is that the residual error must be known a priori. The method, therefore, is nonparametric with respect to the interindividual random effects but requires the intraindividual error to be specified a priori. PK

analyses performed with the NPML approach and reported in the literature have used a residual error model based on drug concentration measurement assay variance (69–72). This seems to be unrealistic. Intraindividual variability, interoccasion variability, and model misspecification often will contribute significantly to the residual error (73, 74). Also, the estimator of the distribution produced by the NPML approach is a point estimator and no results on the accuracy of the estimation are obtained. Consequently, care should be taken in interpreting the results especially when they are obtained from a small sample size. If the NPML approach is used primarily for exploratory analysis to improve the efficiency of subsequent parametric analysis, this may not be much of a problem. The NPML approach is a computationally expensive approach, which may limit the practicality of the approach when the dimension of the parameter space increases. An example of this would be the case of a complex PK model with numerous covariates.

The nonparametric expectation-maximization (NPEM) program of Schumitzky (75), which is similar to the NPML program of Mentre and Mallet (69) computes the nonparametric maximum likelihood using the nonparametric EM algorithm. NPEM has been developed as a segment of the USC*PACK collection of programs (35). The results obtained using NPEM for PPK data analysis are similar to those of the NPML program. NPEM and STS give virtually identical estimates of PPK parameters in the same population when the results of NPEM indicate normal distribution for parameter estimates (76, 77).

10.3.6 Semi-nonparametric (SNP) Maximum Likelihood

Davidian and Gallant (78) introduced the SNP maximum likelihood from econometrics into pharmacokinetics. Like the NPML approach, the SNP maximum likelihood approach provides an estimate of the entire distribution of the interindividual random effects. The SNP maximum likelihood approach maximizes the likelihood over a class of distributions restricted to have a smooth density, instead of maximizing the likelihood over all distribution functions as the NPML method does. This assumption of smoothness is flexible enough to allow heavy-tailed, multimodal, and skewed distributions to be characterized but prevents kinks, jumps, and oscillatory behavior (79). Also, this method relies on maximizing the likelihood of the set of observations of all individuals to estimate the distribution of the random effects. The representation of the probability distribution and calculation of the likelihood are different from the NONMEM and NPML approaches. It has been shown by Gallant and Nychka (80) that the smooth distribution can be presented as an infinite series expansion, and they provide a full mathematical description. The SNP maximum likelihood approach uses a finite number of leading terms resulting from an approximation of the infinite expansion. A single tuning parameter determines the number of terms retained. The density is multivariate normal if the value of this tuning parameter equals zero. The distribution becomes more flexible, the larger the value of the tuning parameter. An important step in the modeling procedure is the selection of an appropriate value of this tuning parameter (78). The density of the random effect parameters is represented by a multivariate normal distribution multiplied by a polynomial. The SNP maximum likelihood approach computes the integral present in the population likelihood by quadrature. This is another useful feature of this approach. This obviates the use of the linearization approximation

to the likelihood used in the NONMEM approach. Unlike the NPML approach, standard errors can be computed for the model parameters and used for inference.

The SNP maximum likelihood approach is implemented in a public domain Fortran program called NLMIX. Experience with this approach is still very limited and only a few simulations have evaluated the ability of the method to reveal multiple modes in the random effects density under conditions likely to be encountered in practice.

A method similar to the SNP maximum likelihood approach was proposed by Fattinger et al. (81) to explore the complete distribution of interindividual effects using the FOCE approach in the NONMEM program. The method uses a monotone nondecreasing spline to transform the normally distributed interindividual random effects. The model for the interindividual random effect model is given as

$$\phi_i = g(\theta, x_i) + sp(\eta_i) \tag{10.22}$$

where $sp(\cdot)$ represents a monotone nondecreasing spline of which the parameters are estimated. Because splines are not multivariate, a different spline is used for each of the elements of η_i. The spline function transformation is very flexible and allows appropriate representations of skewed, heavily tailed, or multimodal distributions.

10.4   SUMMARY

The methods discussed represent different schools of thought regarding the estimation of nonlinear mixed effects models. They vary widely in distributional assumptions, approximations, and utility in routine analysis of clinical trial/observational data. The practical ability to establish PPK models, which are either descriptive or predictive, for their intended purpose provides for a better understanding of the pharmacokinetics of drugs with a view to optimizing therapy/drug development programs. NONMEM is the most widely used software for estimation of population pharmacokinetics. It is worth noting that the increase in computational power has enabled implementation of Bayesian methodologies.

REFERENCES

1. *Guidance for Industry: Population Pharmacokinetics.* U.S. Department of Health and Human Services, Food and Drug Administration, Washington, DC, 1999.
2. E. I. Ette and P. J. Williams, Population pharmacokinetics 1: background, concepts and models. *Ann Pharmacother* **38**:1702–1706 (2004).
3. E. Chatelut, A. V. Boody, B. Peng, et al., Population pharmacokinetics of carboplatin in children. *Clin Pharmacol Ther* **59**:436–443 (1996).
4. L. Collart, T. H. Blaschke, F. Bouch, et al., Potential of population pharmacokinetics to reduce the frequency of blood required for estimating kinetic parameters in neonates. *Dev Pharmacol Ther* **18** (1–2):71–80 (1992).

5. E. Yukawa, H. Mine, S. Higuchi, et al., Digoxin population pharmacokinetics from routine clinical data: role of patient characteristics for estimating dosing regimens. *J Pharm Pharmacol* **44**:761–765 (1992).

6. P. Crome and R. J. Flanagan, Pharmacokinetic studies in elderly people: are they necessary? *Clin Pharmacokinet* **26**(4):243–247 (1994).

7. S. M. Pai, U. A. Shukla, T. H. Grasela, et al., Population pharmacokinetic analysis of didadosine (2′,3′-dideoxyinosine) plasma concentration obtained in phase I clinical trials in patients with AIDS or AIDS-related complex. *J Clin Pharmacol Ther* **32**:164–169 (1995).

8. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Michealis–Menten model: routine clinical data. *J Pharmacokinet Biopharm* **8**:553–571 (1980).

9. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm* **9**:635–651 (1981).

10. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Monoexponential model and routine clinical data. *J Pharmacokinet Biopharm* **11**:303–319 (1983).

11. E. Samara and R. Granne, Role of pharmacokinetics in drug development: a pharmaceutical industry perspective. *Clin Pharmacokinet* **32**(4):294–312 (1997).

12. B. Whiting, A. W. Kelman, and J. Grevel, Population pharmacokinetics: theory and clinical application. *Clin Pharmacokinet* **11**:387–401 (1986).

13. M. O. Karlsson and L. B. Sheiner, The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Biopharm* **21**(6):735–750 (1993).

14. E. I. Ette and P. J. Williams, Population pharmacokinetics. II: Estimation methods. *Ann Pharmacother* **38**:1907–1915 (2004).

15. M. O. Karlsson, S. L. Beal, and L. B. Sheiner, Three new residual error models for population PK/PD analyses. *J Pharmacokinet Biopharm* **23**(6):651–672 (1995).

16. G. Levy and L. E. Hollister, Inter- and intrasubject variations in drug absorption kinetics. *J Pharm Sci* **53**:1446–1452 (1964).

17. E. Martin, W. Moll, P. Schmid, and L. Dettli, Problems and pitfalls in estimating average pharmacokinetic parameters. *Eur J Clin Pharmacol* **26**:595–602 (1984).

18. D. M. Cochetto, W. A. Wargin, and J. W. Crow, Pitfalls and valid approaches to pharmacokinetic analysis of mean concentration data following intravenous administration. *J Pharmacokinet Biopharm* **8**:539–552 (1980).

19. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Michealis–Menten model: routine clinical data. *J Pharmacokinet Biopharm* **8**:553–571 (1980).

20. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm* **9**:635–651 (1981).

21. L. B. Sheiner and S. L. Beal, Evaluation of methods for estimating population pharmacokinetic parameters. I. Monoexponential model and routine clinical data. *J Pharmacokinet Biopharm* **11**:303–319 (1983).

22. L. B. Sheiner and S. L. Beal, Estimation of pooled pharmacokinetic parameters describing populations, in *Kinetic Data Analysis*, L. Endrenyi (Ed.). Plenum Press, New York, 1981, pp. 271–284.

23. H. Fluehler, H. Huber, E. Widmer, and S. Brechbuehler, Experiences in the application of NONMEM to pharmacokinetic data analysis. *Drug Metab Rev* **15**:317–339 (1984).

24. J. B. Dyck, D. L. Haack, L. Azarnoff, L. Vuorilehto, and S. L. Shafer, Computer-controlled infusion of intravenous dexmedetomide hydrochloride in adult human volunteers. *Anesthesiology* **78**:821–828 (1993).

25. S. L. Shafer, J. R. Varvel, N. Aziz, and J. C. Scott, Pharmacokinetics of fentanyl administered by computer-controlled infusion pump. *Anesthesiology* **73**:1091–1102 (1990).

26. L. L. Gustafsson, W. F. Ebling, E. Osaki, S. Harapat, D. R. Stanski, and S. L. Shafer, Plasma concentration clamping in the rat using a computer-controlled infusion pump. *Pharm Res* **9**:800–807 (1992).

27. B. K. Kataria, S. A. Ved, H. F. Nicodemus, D. Lea, M. Y. Dubios, J. W. Mandema, and S. L. Shafer, The pharmacokinetics of propofol in children using three different data analysis approaches. *Anesthesiology* **80**:104–122 (1994).

28. T. D. Egan, H. J. Lemmens, P. Fiset, D. J. Hermann, K. T. Muir, D. R. Stanski, and S. L. Shafer, The pharmacokinetics of the new short-acting opioid remifentanil (GI87084B) in healthy adult male volunteers. *Anesthesiology* **79**:881–892 (1993).

29. B. E. Rodda, Analysis of sets of estimates from pharmacokinetic studies, in *Kinetic Data Analysis*, L. Endrenyi (Ed.). Plenum Press, New York, 1981, pp. 285–297.

30. J. L. Steimer, A. Mallet, J. L. Golmard, and J. F. Boisvieux, Alternative approaches to estimation of population pharmacokinetic parameters: comparison with nonlinear mixed effects model. *Drug Metab Rev* **15**:265–292 (1984).

31. L. B. Sheiner and S. L. Beal, A note on confidence intervals with extended least squares parameter estimates. *J Pharmacokinet Biopharm* **15**:93–98 (1987).

32. D. B. White, C. A. Walawander, Y. Tung, and T. H. Grasela, An evaluation of point and interval estimates in population pharmacokinetics using NONMEM analysis. *J Pharmacokinet Biopharm* **19**:87–112 (1991).

33. G. Prevost, Estimation of a normal probability density function from samples measured with non-negligible and non-constant dispersion. Internal Report 6–77, Adersa-Gerbios, 2 avenue du 1er mai, F-91120 Palaiseau.

34. A. Forrest, C. H. Ballow, D. E. Nix, M. C. Birmingham, and J. J. Schentag, Development of a population pharmacokinetic model and optimal sampling strategy for intravenous ciprofloxacin in seriously ill patients. *Antimicrob Agents Chemother* **37**:1065–1072 (1993).

35. R. W. Jellife, A. Schumitzky, and M. Van Guillder, *User Manual for the Non-parametrric EM Program for Population Modeling, Version 2.17*. Laboratory for Applied Pharmacokinetics, USC School of Medicine, Los Angeles, CA, 1993.

36. F. Mentre and R. Geomeni, A two-step iterative algorithm for estimation in nonlinear mixed-effect models with an evaluation in population pharmacokinetics. *J Biopharm Stat*; **5**:141–158 (1995).

37. *P-PHARM User's Guide, Version 1.3.* SIMED, Créteil, France, 1994.

38. A. Racine-Poon, A Bayesian approach to nonlinear random effects models. *Biometrics* **41**:1015–1023 (1985).

39. A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via EM algorithm. *J R Stat Soc B* **39**:1–38 (1977).

40. A. Racine, A. P. Grieve, H. Fluhler, and A. F. M. Smith, Bayesian methods in practice: experiments in the pharmaceutical industry. *Appl Stat* **35**:93–100 (1986).

41. D. J. Spiegelhalter, A. Thomas, and W. Gilks, *BUGS Manual 0.60*. MRC Biostatistics Unit, Cambridge, UK, 1997.

42. N. G. Best, K. K. Tan, W. R. Gilks, and D. J. Spiegelhalter, Estimation of population pharmacokinetics using the Gibbs sampler. *J Pharmacokinet Biopharm* **23**(4): 407–435 (1995).

43. D. J. Lunn, N. Best, A. Thomas, J. Wakefield, and D. J. Spiegelhalter, Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn* **29**(3):271–307 (2002).

44. L. B. Sheiner, B. Rosenberg, and K. L. Melmon, Modeling of individual pharmacokinetics for computer-aided drug dosage. *Comp Biomed Res* **5**:441–459 (1972).

45. L. B. Sheiner, The population approach to pharmacokinetic data analysis: rationale and standard data analysis methods. *Drug Metab Rev* **15**:153–171 (1984).

46. S. L. Beal, Population pharmacokinetic data and parameter estimation based on their first two statistical moments. *Drug Metab Rev* **15**:173–193 (1984).

47. B. Whiting, A. W. Kelman, and J. Grevel, Population pharmacokinetics: theory and clinical application. *Clin Pharmacokinet* **11**:387–401 (1986).

48. L. B. Sheiner and S. L. Beal, Bayesian individualization of pharmacokinetics: simple implementation and comparison with non-Bayesian methods. *J Pharm Sci* **71**:1344–1348 (1982).

49. E. I. Ette, A. W. Kelman, C. A. Howie, and B. Whiting, Analysis of animal pharmacokinetic data: performance of the one point per animal design. *J Pharmacokinet Biopharm* **23**:551–566 (1995).

50. E. I. Ette, H. Sun, and T. M. Ludden, Balanced designs and longitudinal population pharmacokinetic studies. *J Clin Pharmacol* **38**:417–423 (1998).

51. T. H. Grasela, Jr., E. J. Antal, R. J. Townsend, and R. B. Smith, An evaluation of population pharmacokinetics in therapeutic trials. Part I. Comparison of methodologies. *Clin Pharmacol Ther* **39**:605–612 (1986).

52. L. Collart, T. F. Blashke, F. Boucher, and C. G. Prober, Potential of population pharmacokinetics to reduce the frequency of blood sampling required for estimating kinetic parameters in neonates. *Dev Pharmacol Ther* **18**:71–80 (1992).

53. L. Aarons, J. W. Mandema, and M. Danhof, A population analysis of the pharmacokinetics and pharmacodynamics of midazolam in the rat. *J Pharmacokinet Biopharm* **19**:485–496 (1991).

54. N. Kaniwa, N. Aoyagi, H. Ogata, and M. Ishi, Application of the NONMEM method to evaluation of the bioavailability of drug products. *J Pharm Sci* **79**:1116–1120 (1990).

55. S. M. Pai, U. A. Shukla, T. H. Grasela, et al., Population pharmacokinetic analysis of didanosine (2′,3′-dideoxyinosine) plasma concentration obtained in phase I clinical trials in patients with AIDS or AIDS-related complex. *J Clin Pharmacol* **32**:242–247 (1992).

56. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional population pharmacokinetic studies: implications for pediatric and animal studies. *Clinc Res Regul Affairs* **13**:133–165 (1996).

57. A. Racine, A. P. Grieve, H. Fluhler, and A. F. M. Smith, Bayesian methods in practice: experiences in pharmaceutical industry. *Appl Stat* **35**:1–38 (1986).

58. E. I. Ette, A. W. Kelman, C. A. Howie, and B. Whiting, Influence of interanimal variability on the estimation of population pharmacokinetic parameters in preclinical studies. *Clin Res Reg Affairs* **11**:121–139 (1994).

59. E. I. Ette, C. A. Howie, A. W. Kelman, and B. Whiting, Experimental design and efficient parameter estimation in preclinical pharmacokinetic studies. *Pharm Res* **12**:729–737 (1995).

60. M. O. Karlsson and L. B. Sheiner, The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Biopharm* **21**(6):735–750 (1993).

61. S. L. Beal and L. B. Sheiner, *NONMEM Users Guide—Part VII. Conditional estimation Methods*. University of California, San Francisco, 1992.

62. M. J. Lindstrom and D. M. Bates, Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**:673–687 (1990).

63. *S-Plus* ™. Insightful, Seattle, Washington, 2002.

64. E. F. Vonesh and R. L. Carter, Mixed–effects nonlinear regression for unbalanced repeated measures. *Biometrics* **46**:673–687 (1992).

65. E. F. Vonesh, Nonlinear models for the analysis of longitudinal data. *Stat Med* **11**:1929–1954 (1992).

66. R. D. Wolfinger, Laplace's approximation for nonlinear mixed models. *Biometrika* **80**:791–795 (1993).

67. A. Mallet, A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73**:645–656 (1986).

68. J. L. Steimer, A. Mallet, and F. Mentre, Estimating interindividual pharmacokinetic variability, in *Variability in Drug Therapy: Description, Estimation, and Control*, M. Rowland et al. (Eds.). Raven Press, New York, 1985, pp. 65–111.

69. F. Mentre and A. Mallet, Experiences with NPML—application to dosage individualisation of cyclosporine, gentamicin and zidovudine, in *The Population Approach*, M. Rowland and L. Aarons (Eds.). Commission of European Communities, Luxembourg, 1992, pp. 75–90.

70. A. Mallet, F. Mentre, J. Gilles, A. W. Kelman, A. N. Thomson, S. M. Bryson, and B. Whiting, Handling covariates in population pharmacokinetics with an application to gentamicin. *Biomed Meas Inform Contr* **2**:673–683 (1988).

71. A. Mallet, F. Mentre, J. L. Steimer, and F. Lokiec, Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine. *J Pharmacokinet Biopharm* **16**:529–556 (1988).

72. F. Mentre, A. Mallet, B. Diquet, et al., Population kinetics of AZT in AIDS patients. *Eur J Clin Pharmacol* **36**:230 (1989).

73. J. S. Sidhu, M. Ashton, N. V. Huong, et al., Artemisinin population pharmacokinetics in children and adults with uncomplicated falciparum malaria. *Br J Clin Pharmacol* **45**:347–354 (1998).

74. M. Hossain, E. Wright, R. Baweja, T. M. Ludden, and R. Miller, Nonlimear mixed effects modeling of single dose and multiple dose data for an immediate release (IR) and controlled release (CR) dosage form of alprazolam. *Pharm Res* **14**:309–315 (1997).

75. A. Schumitzky, Nonparametric EM algorithms for estimating prior distributions. *Appl Math Comput* **45**:141–157 (1991).

76. W. F. Dodge, R. W. Jellife, C. J. Richardson, R. A. McCleery, J. A. Hokanson, and W. R. Snodgrass, Gentamicin population pharmacokinetic models for low birth weight infants using a new nonparamaetric method. *Clin Pharmacol Ther* **50**:25–31 (1991).

77. R. W. Jellife, P. Gomis, and A. Schumitzky, A population model of gentamicin made with a new nonparametric EM algorithm. Technical Report 90-4. Laboratory for Applied Pharmcokinetics, USC School of Medicine, Los Angeles, CA, 1990.

78. M. Davidian and A. R. Gallant, Smooth nonparametric maximum likelihood estimation for population pharmacokinetcs, with application to quinidine. *J Pharmacokinet Biopharm* **20**:529–556 (1992).

79. M. Davidian and A. R. Gallant, The nonlinear mixed effects model with a smooth random effects density. Institute of Statistics Mimeo Series No. 2206, North Carolina State University, Raleigh, NC, 1992.

80. A. R. Gallant and D. W. Nychka, Semi-nonparametric maximum likelihood estimation. *Econometrica* **55**:363–390 (1987).

81. K. E. Fattinger, L. B. Sheiner, and D. Verotta, A new method to explore the distribution of interindividual random effects in nonlinear mixed effects models. *Biometrics* **51**:1236–1251 (1995).

Timing and Efficiency in Population Pharmacokinetic / Pharmacodynamic Data Analysis Projects

SIV JÖNSSON and E. NICLAS JONSSON

11.1 INTRODUCTION

This chapter addresses the time requirements and the efficiency of population data analysis projects in drug development, with focus on the rapid delivery of results to the clinical project team and decision makers. Data analysis projects are considered to involve a number of different tasks, with the main milestone being the communication of the results in time for decision making. Significant time savings can be achieved by adopting one of two main approaches: (a) changing the timing of the tasks and/or (b) performing individual tasks more efficiently. How such time savings apply to each task in a data analysis project is discussed, and a best case scenario is presented.

The population pharmacokinetic/pharmacodynamic (PK/PD) modeling approach has been advocated as an essential tool to improve the efficiency and to facilitate decision making in drug development and regulatory assessment (1–8). The experiences reported by the pharmaceutical industry have shown that the use of PK/PD modeling influenced the direction of development programs and was associated with time savings (9–13). The latter is of particular interest since one major criticism of population modeling within industry is that it is time consuming. This criticism is justified: modeling takes time and, unless proper and conscious care is taken to smooth the model-based analysis process, the effort expended on modeling might only result in a tick on the checklist for a regulatory submission package.

Despite the fact that this chapter focuses on the data analysis of data from a single study, it should not be forgotten that the multistudy, multiphase, and multiyear nature of drug development is a modeling asset. In practice, studies have always preceded a current one and it can be assumed that there will always be studies following it. This fact has an impact on the strategy adopted when modeling

in any one project and thus will affect the time efficiency. This means that there is a certain amount of existing knowledge to learn from and on which assumptions for the present analysis can be based. Likewise, the present analysis always has the possibility to impact future studies and decision making, so by keeping potential research questions in mind during the analysis it is possible to improve the long-term value of the present modeling effort.

Another important aspect of the long-term nature of drug development programs is that the "cost" of time is dependent on what drug development phase one is discussing. In early phases of the development, research effort is concentrated on learning about the drug and the biological system in which it operates, whereas in later phases the emphasis is on obtaining unequivocal confirmation that the drug is both safe and efficacious (14). Clearly, later phases will be subject to more stringent deadlines and therefore will have more arduous timing requirements than the earlier phases.

From the point of view of modeling, the early drug development phases are ideal for model characterization; that is, this is the period in which to undertake the main aspects of modeling that require a considerable amount of time and for which it is difficult to estimate the time required. Another less exciting—but nonetheless important—aspect is that it is in the early phases of the drug development program that there is time to set up the documentation and model building infrastructure required for all forthcoming data analysis projects for the candidate drug. This includes a plan for the data analysis, report templates, model building and diagnostic routines, and specifications for the creation of data sets (as explained in greater detail later).

In the later phases of a drug development program, the focus is likely to be on confirmation of the characteristics of the drug rather than on exploration. From the perspective of modeling, there are two scenarios. The first is that modeling was systematically used in earlier phases of the program and, thus, a likely model has been characterized and the documentation and model building infrastructure is already in place. In this instance, the pharmacometricians and the model will be ready to make the transition to the confirmatory model. In essence, the requirement is for the plans for the data analysis to be specified in greater detail, for the procedures to be undertaken to be specified in advance, and for the aims of the data analysis to be well defined. The second scenario is that essentially no modeling was conducted in the earlier phases. In such a situation the modeling project, which may very well be on the critical path to regulatory filing, will need to start from scratch in almost all respects. The model will need to be characterized, model building routines and procedures will need to be set up, and the requirements for the extraction of data from the clinical database must be agreed on with the database programmers—and the list continues. Under this scenario, a tremendous amount of resources and dedication will be necessary to make the model-based analysis efficient enough to deliver results in time to add value to the decision-making process.

In this chapter we focus on pinpointing problems related to keeping a project on time. The chapter is organized in a manner that corresponds to the individual components of a data analysis project, starting with the planning phases, then the actual modeling, and ending with the report-writing phase. We discuss how each component can be made more efficient and/or how its place in the project can be altered to reduce the time between the delivery of the final data set to the com-

munication of the results. It should also be pointed out that, although much of what is discussed applies regardless of what data analysis software is used, our own experience of population modeling has been obtained with NONMEM® (15) and any examples are based on NONMEM.

11.2 SAVING TIME IN A POPULATION MODELING PROJECT

11.2.1 Planning

Planning is crucial if one is to complete a population modeling project as efficiently as possible (16). A population modeling project can be characterized by its goal and the various tasks involved in the project, including their duration. Figure 11.1 shows the time schedule allocated to a project, starting with the definition of the
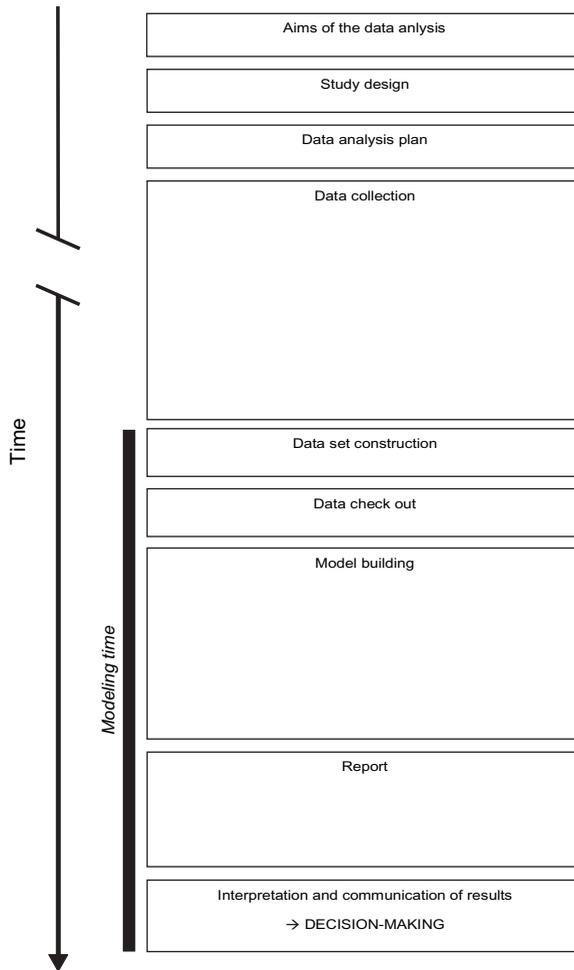
FIGURE 11.1. Assignments in a population modeling project. The solid black line indicates the time period referred to as the *modeling time.*

question to be answered and ending with the final report and the subsequent decision making. The major tasks in between are (naively) lined up in serial fashion. When it comes to the time requirements, the important aspect, in terms of the modeling efficiency as perceived by colleagues in the project team, is the period between the end of the data collection phase and the feedback of the knowledge obtained to the project team and/or the decision makers (this is marked with a solid line in Figure 11.1). We refer to this period as the *modeling time*. If the modeling time is long, then those awaiting the results will perceive the model-based analysis to be less efficient than if it were shorter. However, what is more important than the perception of efficiency is the fact that when the modeling time is short, decision makers can be informed of the outcome of the analysis sooner. In other words, it is the modeling time pharmacometricians should aim at reducing. The overall time required for the clinical study is, of course, also important, but this is usually outside the control of the pharmacometrician.

11.2.2 Approaches

There are basically two approaches to the reduction of the modeling time:

1. Accomplishing tasks at times other than when the modeling is being conducted. In Figure 11.1, the tasks involved in the project are laid out in a sequential manner. However, many of these tasks could be performed at an earlier stage in the project or in parallel with other tasks. The implementation of this change means that the thought that must go into accomplishing these tasks can take place outside the modeling time.
2. Performing tasks more efficiently. When it is not possible to reschedule tasks to conduct them at a more convenient time, it is often possible to reduce the time taken to perform them. For example, the way a task is done can be modified or preparatory work can be conducted in advance, at earlier stages in the project.

In the following we go through the different parts of a population modeling project where we believe significant time savings can be made by applying either of the approaches proposed above.

11.3 THE POPULATION MODELING PROJECT

11.3.1 Aims of the Population Analysis

"If you know where you are going then you know when you get there" is a saying that holds very well in a population modeling project. If there is a clear, measurable goal for the modeling, we will know when to stop the ongoing procedure and enter into the next phase, that is, the communication of the results. The definition of the goal needs to include not only what we want to know but also what we are willing to assume and how certain we need to be. The latter is especially important since it is always possible to refine a model just a little bit more, even if the resultant refinement is of minor practical importance or none at all.

Naturally, the definition of the goal of the modeling will have a bearing on the design of the study and the plan for the analysis as they need to reflect the goals to be met. Writing the plan for the analysis, for example, will be easier and more rapidly achieved if the intended scope of the analysis has been limited and/or well defined. Accordingly, defining the goals properly is critical if one is to end up with a realistic plan for the analysis and is an iterative process necessitating the cooperation of scientists from various disciplines—clinicians, regulators, pharmacometricians, and statisticians (17).

Furthermore, time can be gained by dividing the goals into (a) primary aims that are critical for decision making and should be communicated as quickly as possible to the relevant people, and (b) secondary aims related more to exploration and learning, the results of which will be of more long-term value.

11.3.2 Designing the Study

The design of the study will clearly depend on the overall aims and the mode (modeling or other) of the main analysis. Clearly, the design should allow the main aims to be met with the chosen data analysis methodology.

If it is decided that the data from a clinical trial is to be analyzed using a model-based approach, regardless of whether this is the main mode of analysis or not, the design will need to allow the goal of the modeling to be met. If this is not the case, then the best way to save time is to drop the modeling. Furthermore, if the design leads to low information content, then the results will provide little information. The primary way modeling, and any other mode of analysis, can turn a bad set of data into informative results is by adding information, that is, making more assumptions and adding knowledge that existed prior to the analysis. Although solving the issues related to the informativeness, adding assumptions weakens the results of the analysis and decreases the possible number of new insights. In the extreme case, the final model will only describe what was known about the drug before the study and the actual data will not add anything. Applying all the tools available, for example, optimal design theory (18, 19), informative block randomized design (20), and clinical trial simulations (21), to obtain a design that is powerful and informative from a modeling perspective will not only make the final model adequate but will also lead to a more time-efficient model building process. A more detailed discussion of issues that have an impact on study design is provided in other chapters of this book.

11.3.3 Writing the Plan for the Data Analysis

The *Guidance for Industry: Population Pharmacokinetics* clearly states that any type of population pharmacokinetic (PK) analysis should be defined in a protocol (7). The justification for this is that objectivity should be retained through prespecification of all procedures and methodologies that will be used. In the pharmaceutical industry this protocol usually takes the form of a plan for the analysis, either produced as part of the clinical trial protocol or as a separate document.

Table 11.1 lists some useful headings in a data analysis plan. The specific outline will vary from company to company and/or from project to project. Clearly, there is a great deal of information included in this type of document, so it is necessary to realize that writing it will require a considerable amount of time. On the other

TABLE 11.1 Examples of Useful Items in a Data Analysis Plan

Item	Description
Overall clinical trial objectives	Specification of the overall objectives of the clinical trial
Study design	Provision of the overall design of the study including the specification of drug treatments, procedures for the collection of pharmacokinetic and pharmacodynamic data, and description of the bioanalytical methods
Objectives of population analysis	Specification of the specific objectives of the population analysis
Data	Stating which data are to be included, the data input format required by NONMEM, how to deal with missing data, and how to handle outliers
Data analysis method	Stating the software, model building procedures, model diagnostics, structural model, covariate model, stochastic model, and sensitivity analysis to be used and how the evaluation of the model is to be conducted
Presentation of results	Laying out which results are to be presented in a tabular format, the figures to be presented, the appendixes that need to be included, and any electronic files that should be prepared
Time plan	Providing a time frame for the preliminary model building and the production of clean file, clean data sets, the analysis, and the report-writing

hand, a well-written and well-considered analysis plan will be a great asset to the pharmacometrician once the clock starts to tick. If the actual modeling is to be done by an external consultant, the analysis plan takes on the role of a "product specification" when stating how the analysis is to be performed.

Another important point to remember is that the analysis plan should be finalized before the data from the study becomes accessible. On the other hand, it does not need to have been finalized before the actual study (the data collection) starts, meaning that some sections and many details can be added while the study is running. However, since the analysis plan describes how the analysis should be performed and as this, in turn, depends on the goals of the analysis, it is necessary to start thinking about the intended form of the analysis and start considering the plan at the beginning of the project. At the start of the data collection the analysis plan need only be sufficiently detailed to ensure that goals of the analysis can be met. The plan can then be finalized some time later, but it must be fully specified before the start of the data analysis.

Writing the analysis plan forces the pharmacometrician and the project team to make decisions about issues related to the modeling that would result in inefficient use of time during the analysis unless they are resolved. Examples of such issues are the identification of appropriate diagnostic plots (see Chapter 7), covariates to include in the analysis and the handling of correlated covariates (22–24), how to handle missing data, what to do about outliers (6, 7), criteria to be used for model discrimination (25–27), and how to communicate the results immediately after finalization of the analysis. The most important aspect of this is that a large part of

the *thinking* is done before the modeling commences so that, when an issue arises during the modeling, it is just a matter of consulting the analysis plan to determine the appropriate course of action.

13.3.4 Data Collection

Clearly, the collected data needs to be of good quality. This is not only important from a strict quality point of view, but also from the perspective of modeling: outright errors will take time to sort out, partly because it must first be recognized that there is an error, and partly because the source data (the clinical database) has to be checked and a new data set created. "Non-errors," such as per-protocol time specifications of observations and dosing events instead of the actual time points, will reduce the quality of parameter estimates and may lead to longer run times (28). Thus, when one is considering optimizing the overall time required for a modeling project, great effort should be put into the prevention of errors and into initializing the data processing and the cleanup of data as soon as observations have been made; for example, real-time data assembly should be implemented when possible (29, 30). To a great extent, occurrence of errors may be prevented by education of all personnel involved in data collection.

From the perspective of the time required for modeling, it is apparent that a very important aspect of the data collection phase is ensuring that the pharmacometrician takes the time to prepare for the modeling. This preparatory work should include finalization of the data analysis plan, preparation of model building procedures, and construction of a template or templates for the report. In this way, the data collection phase can shorten the time required for modeling.

11.3.5 Data Set Construction

The construction of data sets for population modeling may be considered a trivial issue on first consideration, but in-depth reflection reveals this to be one of the main practical challenges to the pharmacometrician. Underestimation of the time this requires is a common mistake. Regardless of which software is to be used for the analysis, the data must be organized in a way that makes it interpretable by the software employed. The complexity of organizing the data set depends on the amount of data and whether specific information is required by the chosen software packages.

Smooth construction of the data set can be accomplished by detailed planning and frequent interaction with database personnel/programmers. If possible, the personnel working with the database should receive special training in construction of, for example, NONMEM data sets. For NONMEM, the organization of a data set involves keeping each individual record, that is, all data related to one patient, in chronological order, including information related to the dosage history, which may necessitate additional data items. Furthermore, in NONMEM, the data structure defines part of the model by inclusion of certain data items specific to NONMEM, for example, RATE and CMT. The master data file, obtained from the clinical database, should be as complete as possible with respect to such data items to facilitate the model building. This includes data items necessary for alternative model building paths, for example, a RATE column (with the appropriate

entries), even if the drug is expected to exhibit first-order absorption. To avoid misunderstandings, the database staff should be provided with a specification of the format of the data sets, that is, a template of the data set, including thorough explanations for each data item. This information is naturally included in the analysis plan, meaning that personnel involved in data set construction will be involved from the very start.

The probability of successful retrieval of data from the database and construction of the analysis data set can be enhanced by training in advance with data similar to that to be obtained or with dummy data. Such a procedure is invaluable to identify steps where errors could occur and data items that might be problematic, examples being time-varying covariates or complicated dosing histories. This has been reported to be a very cost-effective procedure (29). The training procedure should also involve data checkout, which leads to the early development of informative data checkout plots and scripts. While data checkout might be perceived to be time consuming at the time it is being conducted, it represents a significant time saving later on because errors are identified at an early stage before they have influenced the model development. A few erroneous data records may cause great difficulties in the model development. Hence, by applying such training procedures, several routines can be prepared for use on the final data set, for example, programming for data assembly, summarizing data and data checkout, and recognition of useful plots for data checkout (for some specific examples, see Chapter 7). When the data from the study is considered to be clean (a clean file is defined at the time point when the database is locked after controlling, completing, and correcting the data, and statistical analysis can be started), creation of the real data analysis file, that is, the file with the real observations from the clinical study, should be a simple matter of pressing a button because all the preliminary work has been conducted beforehand.

During a population analysis certain situations will always require some modification of the data set; for example, the inclusion of data items defining the model or construction of a data set including only a subset of the data. These types of data set modifications are usually the responsibility of the person doing the modeling, and a general piece of good advice is to use one master data set, that is, the one obtained from the clinical database, including all types of measurements and data items, and to base all other data sets on this file. In this way, all data sets have the same origin and are produced in parallel (in contrast to being produced in a sequential manner with one sub-data set being based on another). Using such a parallel approach makes it easy to reproduce data sets and to correct errors found in the data at a late stage in the model development. Furthermore, continuously keeping track of the different data sets during the model development is a must for all pharmacometricians because it will result in substantial time savings in the report-writing phase, not to mention the implications for reproducibility and auditing.

11.3.6 Model Development

When the data has reached a clean file status, time pressure increases for the modeling of the data. This phase of the project involves the characterization of the model as well as its evaluation. There are a number of different approaches, which will be described next, that can be employed to save time or to be more efficient in this

phase. Some of these may need to be given considerable attention early on in the project if it is to be possible to realize them in practice.

11.3.6.1 Early Access to Data

The time required for model development during the time period in which the modeling takes place can be reduced if the model building has already been started before a clean file has been obtained or even before the complete study (i.e., the complete data collection) has been finished.

One possibility is to start modeling before the data is unblinded, that is, before the actual treatment given to each patient has been disclosed (16). To allow the population pharmacometrician access to the data before unblinding demands implementation of standard operating procedures (SOPs), ensuring that the data are blind to other members of the project team. Such SOPs are sometimes implemented for the bioanalytical department, and similar routines can be developed to allow the pharmacometrician access to the data at an early stage too. The possibilities for this may vary greatly from one company to another. Working with the data before it has been unblinded may imply that the data is not clean. However, the major components in the model, for example, the number of compartments in a PK analysis or the major covariate effects, can all be identified even if the data file is based on unclean data. If this approach is adopted, one should be liberal in the exclusion of suspicious outlying data points. This is not as strange as it might sound—it can be expected that some of the information in the data set is wrong, and the most likely points to be erroneous are the outlying data points; there is no reason why a model should be built incorporating data file errors. Once the database has been cleaned, the important modeling steps should be conducted again with all the data included.

11.3.6.2 Increasing Efficiency

The time required for model development can also be decreased by performing parts of the model development in an automated fashion. The total run time will of course be the same but the time between runs will be minimal. Apart from saving time, automation also decreases the risk of the user introducing errors. Examples of modeling tasks that are suitable for automation are those that are repetitive in nature, such as stepwise covariate model building (31) and tasks performed for validation (e.g., likelihood ratio profiling and bootstrapping of the data). If the computational infrastructure allows for parallelization, automation of runs can be even more efficient. At present, there are some generally available routines for automation of such tasks (32, 33).

11.3.6.3 Handling Long Run Times

The run time of a specific model depends on the complexity of the model, the converging algorithm, the chosen software, and the amount of data. In many situations, it is the run time that is the rate-limiting step of model building. The following are some approaches that can be adopted to decrease the duration of or to enable one to cope with long run times.

- *Choosing the Right Estimation Method.* In NONMEM it is the first-order conditional estimation method with or without interaction (FOCE-INTER/FOCE)

that should usually be the estimation method of choice (34). This method is prone to long run times, especially for complex models implemented with differential equations. In many situations, though, the simpler first-order (FO) method in NONMEM gives qualitatively the same results and should therefore be considered during the model development. It will be necessary occasionally to run the latest version of the model using the FOCE-INTER/FOCE method to ensure that it does not give very different results (these being suitable for overnight batch runs). The FOCE-INTER/FOCE method should also be used for specific key runs at stages where decisions affecting the path of the model development are made.

- *Avoid Unnecessary Estimation of Standard Errors.* The estimation of the standard errors is related to the estimation through the inclusion of $COVARI-ANCE in NONMEM. In many situations, though, the calculation of standard errors is as time consuming as the parameter estimation. Hence, under such conditions, the $COVARIANCE step can be used selectively during the development of the model: for example, by only including $COVARIANCE in key runs, that is, the runs to be used in decisions about how to proceed in the model development and for which one needs to be certain that a global minimum has been attained.

- *Use Good Initial Parameter Estimates.* The choice of initial estimates will influence run times. Initial estimates that are far from the final parameter estimates tend to give longer run times than if the initial estimates are close to the final values of the parameters. It is general practice to update the initial estimates throughout the modeling to the best guess initial estimates to avoid long run times and to minimize the risk of obtaining local minima. In most situations the best guess estimates are those obtained from the previous model upon which the subsequent model is based. It is a good habit to compare the initial value of the objective function with the one obtained in the previous model to make sure that the new search starts at a point close to the previous global minimum.

- *Terminate Unsuccessful Runs Early.* The gradient vector, which is available in the output of a run in NONMEM, is helpful for monitoring runs. During a search, the values of the gradients should decrease to low values, but no gradient should be zero at the successful termination of a run. A zero gradient in the initial (0th) iteration is a signal that the model has been coded incorrectly. Such runs should be terminated directly and the model revised. A gradient that becomes zero during the search indicates that there is a problem with the model. Gradients that become zero during a run will rarely take on a nonzero value again. Consider terminating such runs and revise the model.

- *Reducing the Data.* Certain types of data measurements may be problematic due to the volume of data, for example, when observations are collected frequently as in the case of ambulatory blood pressure measurements and EEG measurements. One approach in such situations is to exclude data during the model development. This can be done in a completely random fashion or by applying a stratified random selection strategy (see Chapter 7). The rationale for such data dilution is that the qualitative characterization of the model does not require all of the data in the majority of cases. The final model should, of

course, be confirmed using all of the data available. The final characterization of the parameters should also be based on all the data. This is potentially a controversial approach, but the consequences for model building and any eventual time savings can be assessed on simulated data before the modeling commences, and it will not, therefore, add any to the length of time the project takes.

11.3.6.4 *Use the Time Between Runs Efficiently*

In situations where the run times are short or moderate, the rate-limiting step is often the analyst rather than the computer: that is, more time is spent evaluating the runs than running them. Although evaluation is a better way of spending the modeling time than waiting for the computer, there are ways of making this process more efficient.

Identifying useful diagnostics, finding ways to produce them efficiently, and, most important of all, using them consistently throughout the analysis are all ways to increase efficiency. Various types of plots and certain numerical procedures can be useful for diagnostics (see Chapter 7). A suggestion of the most useful forms of diagnostics can already be made when the plan is written for the data analysis, although, of course, in the light of the data, they may have to be refined, but with some experience the need for such revision is likely to be surprisingly small.

A great deal of time can be spent on producing a particular plot. If the same amount of time and effort needs to be expended for each run, it will not only be inefficient, but there will be a high risk that the plot will not be used consistently (see below). A better approach is to automate the construction of the graph using graphics software packages allowing plots to be scripted (e.g., S-Plus, Insightful Corporation, Seattle, WA; and R, `www.r-project.org`). This will initially take some time, but once the script has been written it is just a matter of invoking it after each run to obtain the same type of plot based on the new model fit. In other software packages it is possible to create templates that can be used to the same effect.

Using the chosen diagnostics methods consistently is also important, partly because the diagnostics provide a way of interpreting and evaluating the run, but also because the increasing familiarity with a certain set of diagnostics during the course of an analysis will make appreciation of changes in a model much easier and quicker.

Once the diagnostics have been produced it is necessary to decide on the next step in the model development. Sometimes this is quite easy, for example, when going from a one-compartment to a two-compartment model, but other decisions can be more problematic unless there are well-defined criteria or established rules. Handling outlying data points is one example: unless the analysis plan clearly states what should constitute an outlying data point, a great deal of time can be spent pondering this question or making test runs to investigate the effect of different choices. The important message here is that the thinking required to resolve such issues does not need to be done during the modeling time.

Keeping a record of all runs is mandatory from a documentation point of view. What should not be overlooked, though, is that such a record can also be a great time saver: with a good record of the runs it will not be necessary to waste time trying to identify a particular model fit and the records will also facilitate report-writing. Even the run record construction can be automated, or at least

semiautomated, through the use of external software, for example, Xpose (35) or Perspective (Cognigen Corporation, Buffalo, NY).

Prespecifiable and housekeeping type runs can be performed efficiently out of working hours using scripts or batch runs. This will save modeling time since the pharmacometrician can spend the working hours to perform those runs that need supervision.

11.3.7   Communication and Interpretation of Results

When the final model has been settled, the main conclusion should be communicated to the relevant people as soon as possible. There is no need to wait until the report has been finalized to communicate some results: if the results from a population analysis are to have an impact on drug development decisions, they need to be communicated (see Chapter 37 also).

11.3.8   Report

The final task once the final population model has been established is to produce a written report. Writing the report usually takes a significant amount of time, as does also the ultimate finalization of the report, especially if the document needs to be approved at several management levels. However, the timing of the finalization of the report is not generally as critical as the immediate communication of the results within the project team. Nevertheless, it is possible to make this ultimate report-writing more efficient by ensuring that the population analysis was well prepared and documented as the model was developed. Much of the standard text in the report can be written before the actual modeling starts and, to a great extent, the analysis plan can serve as a basis for the section of the report dealing with the methods. Furthermore, the structure of the tables in which the demographic information, omitted data, and modeling results are to be presented can be determined and prepared in advance. Ensuring that documentation is kept continuously during the analysis (e.g., comprising run records and data file documentation) is a tremendous help during report-writing. Finally, existing reports from other projects can also serve as templates regardless of whether or not they consider the same compounds because similar issues are likely to be discussed, providing a ready-made framework for the presentation of the data.

11.4   REVISED TIME PLAN

We have identified several possibilities for reducing the modeling time in a population modeling project. Figure 11.2 presents a summary of our discussion and proposals. Again, as in Figure 11.1, there is a line indicating the modeling time but now the principles of moving tasks upstream and parallelizing tasks are illustrated. For example, a comparison with Figure 11.1 reveals that during the period in which the data collection is taking place, several tasks are being prepared for the upcoming data analysis. Figure 11.2 also indicates a possible best case scenario, namely, that preliminary model building can start prior to obtaining a clean file. The time required for the data set construction, checking out the data, and the model
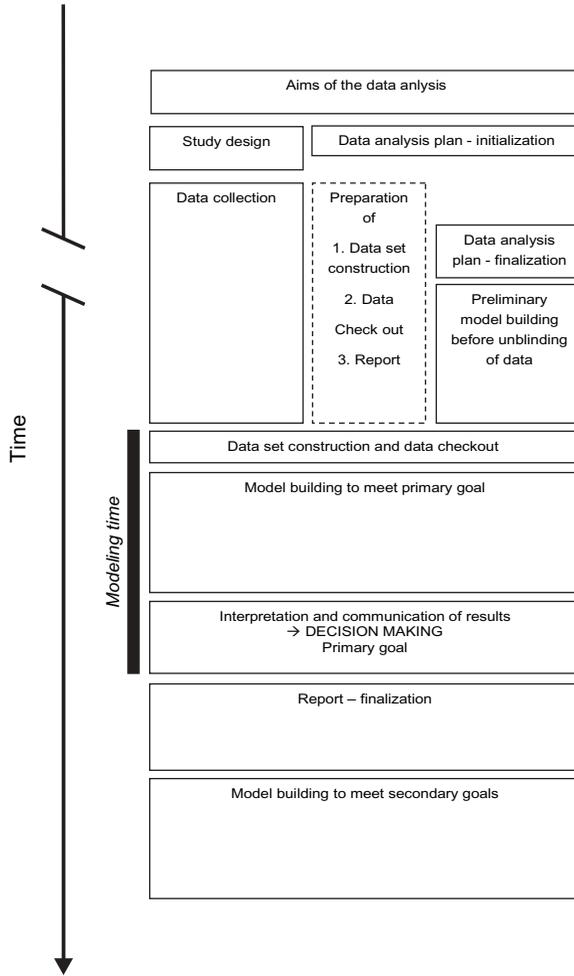
FIGURE 11.2.   Revised time plan for a population modeling project. The box with dashed lines indicates tasks performed intermittently during this time period. The solid black line indicates the time period referred to as the *modeling time.*

building for the clean data set have clearly been reduced (although the length of the boxes should not be considered to be representative of an exact amount of time); this has been accomplished by moving tasks as well as by conducting them more efficiently, as described in this chapter. In addition, Figure 11.2 incorporates the idea that goals can be divided into primary and secondary goals, where fulfilling the primary ones is necessary for decision making, and the secondary ones aim at exploring and learning from the data.

   All the proposed revisions of the project plan contribute to the most important point, namely, that the time taken to deliver the results to the project group has been considerably shortened, thereby increasing the possibility of having an impact on the decision making. The time plans in Figures 11.1 and 11.2 represent two extremes: completely sequential performance of tasks and maximal

parallelization of tasks. No project, even with the ultimate single-mindedness, will adhere to Figure 11.1 without some deviation because some tasks can obviously be prepared in advance or conducted more efficiently. Similarly, however, from a logistical point of view, it is hard to achieve the well-prepared and well-organized ideal represented in Figure 11.2. The overall concept, though, is that with good planning and foresight it should be possible to more nearly attain the plan revealed in Figure 11.2 and thereby reduce modeling time.

11.5   SUMMARY

We do not imply that the proposals given here are the only ways to improve the efficiency in a population modeling project. Indeed, all population modeling projects are different; so some of the ideas presented in this chapter may not always be applicable and, therefore, suitable solutions need to be set up for each project. The way to proceed in each specific case will be dependent on the resources available and the organization of the company. Furthermore, the applicability of our ideas is also dependent on whether the data analysis project is in the initial, intermediate, or late phases of the drug development program. However, hopefully this chapter has given the reader some concrete ideas on how efficiency can be improved and some thoughts on how to make PK/PD modeling an efficient aid in drug development decision making. Planning and interdisciplinary communication within the team may not resolve all problems and obstacles during a modeling project, but they are definitely good foundations for success in keeping within the time constraints of a modeling project.

REFERENCES

 1. L. Sheiner and J. Wakefield, Population modelling in drug development. *Stat Methods Med Res* **8**:183–193 (1999).

 2. S. Vozeh, J. L. Steimer, M. Rowland, et al., The use of population pharmacokinetics in drug development. *Clin Pharmacokinet* **30**:81–93 (1996).

 3. E. Samara and R. Granneman, Role of population pharmacokinetics in drug development. A pharmaceutical industry perspective. *Clin Pharmacokinet* **32**:294–312 (1997).

 4. S. E. Tett, N. H. G. Holford, and A. J. McLachlan, Population pharmacokinetics and pharmacodynamics: an underutilized resource. *Drug Inf J* **32**:693–710 (1998).

 5. S. G. Machado, R. Miller, and C. Hu, A regulatory perspective on pharmacokinetic/pharmacodynamic modeling. *Stat Methods Med Res* **8**:217–245 (1999).

 6. H. Sun, E. O. Fadiran, C. D. Jones, et al., Population pharmacokinetics. A regulatory perspective. *Clin Pharmacokinet* **37**:41–58 (1999).

 7. *Guidance for Industry. Population Pharmacokinetics*, 1999. Available from `http://www.fda.gov/cder/guidance/` (cited March 7, 2004).

 8. L. B. Sheiner and J. L. Steimer, Pharmacokinetic/pharmacodynamic modeling in drug development. *Ann Rev Pharmacol Toxicol* **40**:67–95 (2000).

 9. P. Chaikin, G. R. Rhodes, R. Bruno, S. Rohatagi, and C. Natarajan, Pharmacokinetics/pharmacodynamics in drug development: an industrial perspective. *J Clin Pharmacol* **40**:1428–1438 (2000).

10. B. G. Reigner, P. E. Williams, I. H. Patel, J. L. Steimer, C. Peck, and P. van Brummelen, An evaluation of the integration of pharmacokinetic and pharmacodynamic principles in clinical drug development. Experience within Hoffmann La Roche. *Clin Pharmacokinet* **33**:142–152 (1997).

11. R. Gieschke and J. L. Steimer, Pharmacometrics: modelling and simulation tools to improve decision making in clinical drug development. *Eur J Drug Metab Pharmacokinet* **25**:49–58 (2000).

12. S. C. Olson, H. Bockbrader, R. A. Boyd, et al., Impact of population pharmacokinetic–pharmacodynamic analyses on the drug development process: experience at Parke-Davis. *Clin Pharmacokinet* **38**:449–459 (2000).

13. K. S. Blesch, R. Gieschke, Y. Tsukamoto, B. G. Reigner, H. U. Burger, and J. L. Steimer, Clinical pharmacokinetic/pharmacodynamic and physiologically based pharmacokinetic modeling in new drug development: the capecitabine experience. *Invest New Drugs* **21**:195–223 (2003).

14. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

15. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. GloboMax, Inc., Hanover, MD, 1989–1998.

16. F. Rombout, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, M. Danhof, et al. (Eds.). European Commission, Directorate—General Science, Research and Development, Brussels, 1997, pp. 183–193.

17. J. V. Gobburu and P. J. Marroum, Utilisation of pharmacokinetic–pharmacodynamic modelling and simulation in regulatory decision-making." *Clin Pharmacokinet* **40**:883–892 (2001).

18. S. Retout, F. Mentre, and R. Bruno, Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Stat Med* **21**:2623–2639 (2002).

19. F. Mentre, C. Dubruc, and J. P. Thenot, Population pharmacokinetic analysis and optimization of the experimental design for mizolastine solution in children. *J Pharmacokinet Pharmacodyn* **28**:299–319 (2001).

20. E. I. Ette, H Sun, and T. M. Ludden, Design of population pharmacokinetic studies. *Proc Am Stat Assoc (Biopharmaceut Sec)*: 487–492 (1994).

21. N. H. Holford, H. C. Kimko, J. P. Monteleone, and C. C. Peck, Simulation of clinical trials. *Annu Rev Pharmacol Toxicol* **40**:209–234 (2000).

22. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

23. J. Ribbing and E. N. Jonsson, Power, selection bias and predictive performance of the population pharmacokinetic covariate model. *J Pharmacokinet Pharmacodyn* **31**:109–134 (2004).

24. P. L. Bonate, The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharm Res* **16**:709–717 (1999).

25. U. Wahlby, M. R. Bouw, E. N. Jonsson, and M. O. Karlsson, Assessment of type I error rates for the statistical sub-model in NONMEM. *J Pharmacokinet Pharmacodyn* **29**:251–269 (2002).

26. U. Wahlby, E. N. Jonsson, and M. O. Karlsson, Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn* **28**:231–252 (2001).

27. U. Wahlby, K. Matolcsi, M. O. Karlsson, and E. N. Jonsson, Evaluation of type I error rates when modeling ordered categorical data in NONMEM. *J Pharmacokinet Pharmacodyn* **31**:61–74 (2004).

28. E. J. Antal, T. H. Grasela, Jr., and R. B. Smith, An evaluation of population pharmacokinetics in therapeutic trials. Part III. Prospective data collection versus retrospective data assembly. *Clin Pharmacol Ther* **46**:552–559 (1989).

29. B. Fotteler, K. Jorga, and J. L. Steimer, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, M. Danhof, et al. (Eds.). European Comission, Directorate—General Science, Research and Development, Brussels, 1997, pp. 397–405.

30. T. H. Grasela, E. J. Antal, J. Fielder-Kelley, et al., An automated drug concentration screening and quality assurance program for clinical trials. *Drug Inf J* **33**:273–279 (1999).

31. E. N. Jonsson and M. O. Karlsson, Automated covariate model building within NONMEM. *Pharm Res* **15**:1463–1468 (1998).

32. L. Lindbom, J. Ribbing, and E. N. Jonsson, Perl-speaks-NONMEM (PsN)—a Perl module for NONMEM related programming. *Comput Methods Programs Biomed* **75**:85–94 (2004).

33. N. Holford, *Wings for NONMEM*, 2003. Available from `http://wfn.sourceforge.net/` (cited October 17, 2004).

34. S. L. Beal and L. B. Sheiner, *NONMEM Users Guide—Part VII: Conditional Estimation Methods*. GloboMax, Inc., Hanover, MD, 1992–1998.

35. E. N. Jonsson, and M. O. Karlsson, Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* **58**:51–64 (1999).

Designing Population Pharmacokinetic Studies for Efficient Parameter Estimation

ENE I. ETTE and AMIT ROY

12.1   INTRODUCTION

Most population pharmacokinetic (PPK) studies are performed as adjuncts to clinical trials. The reason for the population approach is to estimate the typical pharmacokinetics in a population—the interindividual variability, intraindividual/interoccasion variability, and residual variability—and to characterize subpopulation differences (based on subject covariates) of the measured pharmacokinetic (PK) responses. Given the adjunct nature of the studies, there is the constant challenge to make the most from a PPK study performed within the context of the clinical trial without jeopardizing the primary objective of the trial. Pragmatic considerations dictate that data be collected under less stringent and restrictive design conditions. The quality of the PPK parameter estimates is a function of experimental design, and the major goal of most PPK studies is the efficient (precise and accurate) estimation of PPK parameters. Effective use of the population approach demands that consideration should be given to how studies can be designed to obtain efficient estimates of population parameters of the pharmacokinetics of a drug.

Issues encountered in the design of PPK studies, the importance of simulation in evaluating study designs, reliability and robustness of parameter estimates, information theory and sampling design, a simulation study comparison of informative block randomized and population information matrix designs, number of samples per subject, sample size and study power, and the determination of the power of a PPK study for characterizing drug–drug interactions are discussed in the subsequent sections.

12.2   ISSUES IN THE DESIGN OF PPK STUDIES

Establishing the structural model describing the pharmacokinetics of the drug in preliminary studies is useful. This basic information is a prerequisite in designing a sparse sampling scheme that will provide adequate information to estimate the parameters in the model. A suboptimal sparse sampling scheme may necessitate the use of a simpler PK model than the model that best described more intensively sampled PK data. The use of a simpler model is sometimes desirable. For instance, if a feature of the more complex model is not clinically relevant, it may be ignored without having an impact on the overall objective of the modeling exercise. An example is a drug such as gentamicin, with a long terminal half-life that can be characterized with three-compartment model pharmacokinetics; but for practical purposes a one-compartment model is used to describe its pharmacokinetics. "When properly performed, population pharmacokinetic studies in patients combined with suitable mathematical/statistical analysis (e.g., using nonlinear mixed-effects modeling) is a valid approach and, on some occasions, an alternative to extensive studies" (1). Thus, the proper performance of PPK studies in terms of study design considerations is of considerable significance. Because of the sparseness of samples obtained from subjects in a PPK study, the choice of a design that will yield efficient parameter estimates and address study objectives is of utmost importance. Population designs consist of a set of individual designs to be performed in groups of subjects.

Individual designs are composed of one or several sample times supplied by each subject. The quality of PPK parameter estimates are a function of experimental design, and a major goal of most PPK studies is the precise and accurate estimation of PPK parameters. The design factors that affect the quality of parameter estimates are the arrangement of concentrations in time, the number of drug concentrations measured per subject, and the number of subjects. To a certain extent these factors can be controlled by the investigator in a prospective PPK study. In this chapter we limit our discussion to designing prospective PPK studies. Addressing the optimization of these design factors has resulted in a number of publications by several authors over the last two decades, many of which are simulation based (2–16). Thus, the importance of simulation in the design of PPK studies is first discussed. This is followed by a discussion of the factors that can be optimized for the efficient design of a PPK study with an example showing the application of these concepts.

12.3   IMPORTANCE OF SIMULATION IN EVALUATING STUDY DESIGNS

The importance of simulation in evaluating study designs was succinctly addressed in the US Food and Drug Administration's *Guidance for Industry: Population Pharmacokinetics* (1): "Simulation has been employed as a tool to investigate the performance of various sampling designs employed in population pharmacokinetic studies. Shortcomings in study design result in the collection of uninformative data. Simulation of a planned study offers a powerful tool for evaluating and understanding the consequences of different study designs. It can reveal the effect of input variables and assumptions on the results of a planned population

pharmacokinetic study. Simulation, therefore, is a useful tool to provide convincing objective evidence why a proposed study design and analysis is preferred to other competing designs." However, simulating the studies is not sufficient. It is important to evaluate the parameter estimates from a simulation study for reliability and robustness.

12.3.1 Reliability and Robustness of Parameter Estimates

Numerical methods used to fit experimental data should, ideally, give parameter estimates that are unbiased with reliable estimates of precision. Therefore, determining the reliability of parameter estimates from simulated PPK studies is an absolute necessity since it may affect study outcome. Not only should bias and precision associated with parameter estimation be determined but also the confidence with which these parameters are estimated should be examined. Confidence interval estimates are a function of bias, standard error of parameter estimates, and the distribution of parameter estimates. Use of an informative design can have a significant impact on increasing precision. Paying attention to these measures of parameter estimation efficiency is critical to a simulation study outcome (6, 7).

Simulation is useful for evaluating the merits of competing study designs (1–16). Competing study design should be evaluated for power, efficiency, and robustness. In evaluating the power of a study with a particular design, the ability to reject a null hypothesis or to estimate a parameter for a subpopulation such as drug clearance should be examined.

It is also important to evaluate the quality of the results of a simulated population PK study for robustness. Robustness addresses the question: "If my assumptions underlying the study design are wrong, am I still able to meet the objectives of the research project?" Evaluation for robustness may be approached with sensitivity analysis. Evidence of robustness renders the results reasonable and independent of the analyst.

12.4 INFORMATION THEORY AND SAMPLING DESIGN/SAMPLE LOCATION

Most population pharmacokinetic studies are performed as adjuncts to clinical trials. There is the constant challenge to make the most from a population PK study performed within the context of the clinical trial without jeopardizing the primary objective of the trial. There is great interest in study designs that can reduce the total cost of longitudinal PPK studies, without compromising the efficiency of such designs (2–16). In practice, sampling design used in longitudinal PPK studies are often unbalanced. Also, the number of individuals to be studied and the individual designs to be performed when there is a maximum cost (i.e., maximum number of samples) to contend with have not been clearly defined in the literature.

Over the years a lot of work has been published on the design of PPK studies. These have ranged from the empirical approaches (2–4, 6, 7, 9, 12) to those based on information theory (5, 8, 10, 11, 13–16).

12.4.1 Empirical Approaches

The application of the empirical approaches yielded interesting and informative findings. For instance, Al-Banna et al. (3) found that the accuracy and precision of random effect parameter estimates from PPK studies improved dramatically when the number of sampling time points for each subject was increased by a single observation beyond the minimum number of 2 required to estimate the individual parameters in the open one-compartment intravenous (IV) bolus model they examined. They examined several three-sample point designs in which the first and the second time points were fixed, while the third time point was varied. They found that the exact location of the third time point was not critical to parameter estimation.

In addition, Hashimoto and Sheiner (2) examined the effect of population pharmacokinetic/pharmacodynamic (PK/PD) sampling design on the accuracy and precision of population PD parameter estimates of an E_{max} model. Although this was not a PPK study but a PPK/PD study, the findings of the study provided useful information in the design of PPK studies. In the Hashimoto–Sheiner study, the PD parameters were estimated by two alternative methods: (a) simultaneous fit of PK and PD data using a population model and (b) sequential fit of PK and PD data, in which the PK data were fit individually, followed by a population fit of the PD data using individual PK parameter estimates. They also examined the effect of PK model misspecification on PD model parameter estimates by using a one-compartment model to fit data from two-compartment models. They found that even a small number of PK observations, suboptimally sampled, resulted in marked improvement in the estimates of population PD parameters. For a given total number of samples, more efficient PK parameter estimates were obtained with designs having fewer samples per subject, but greater number of subjects. They also found that the simultaneous population PK/PD modeling method was more robust to PK model misspecification than the sequential method, and that the robustness of parameter estimates with respect to model misspecification improved when sampling times were selected at random.

The importance of sample location in population pharmacokinetics was further investigated via simulation by Ette (4) and Ette et al. (6) in the one sample/subject situation using the two time point design with a one-compartment IV bolus model. With the first time point sampled as early as possible and the second time point varied between approximately one and three terminal elimination half-lives, it was found that location of the second time point between 1.4 and 3.0 times the half-life of the drug produced efficient estimates of model parameters. It was concluded that locating the second sample point at ≥1.4 times the drug's elimination half-life provided information for efficient estimation of clearance. This work was further extended to three and four time point designs using the same model parameters used in the two-sample design (6). For the three time point design, the first and second time (i.e., last time) points were fixed while the location of the third (middle) time point was varied. In the case of the four time point design, the first (i.e., located as early as possible) and the second (last) time points were fixed. The second time point was located at approximately three times the elimination half-life of the drug. In addition, the third time point was fixed at approximately one-third the elimination half-life of the drug. The fourth time point was varied from 0.7 to 2.5 times the half-life of the drug. It was concluded that the exact locations of the third and

fourth time points for the three and four time point designs, respectively, were not critical to the overall efficiency of parameter estimation. Thus, the work of Ette (4) and Ette et al. (6) were the prelude to the application of information theory to the design of PPK studies and, in particular, sampling design.

12.4.2 Information Theory-Based Approaches

The most widely accepted theoretical approach of determining optimal sampling times for PK studies is based on the Fisher information matrix, the elements of which are the negative of the expected values of the second-order partial derivatives of the log likelihood (5). The theoretical underpinning of this approach is the Rao–Cramer inequality, which states that the inverse of the Fisher information matrix is the lower bound of the variance–covariance matrix of any unbiased estimator. A commonly used criterion for determining optimal sampling times is maximization of the determinant of the Fisher information matrix, which is known as D-optimality criterion. The determinant is a natural optimality criterion choice as it is a scalar valued measure of the magnitude of a matrix, and it is therefore an overall measure of the information about the parameters. The designs proposed by D-optimization are independent of the selection or transformation of the model parameters. It should be noted that D-optimality criterion give equal weight to all parameters in the Fisher information matrix.

The benefits of using D-optimality to obtain measurements at certain key time points that contain the maximum PK information about model parameters have been discussed (17–19). Information theory suggests that at least two sampling times are needed in a single-dose study for the estimation of individual clearance and volume of a one-compartment model following IV dose administration (20). Using Monte Carlo simulation, D'Argenio (17) found that a repeating p-point design (where p is the number of parameters in a model) led to a reduction in parameter estimate uncertainty when data were collected at optimal sampling times. This algorithm was implemented in the SAMPLE component of the ADAPT II software (21) and requires good prior estimates of the PK parameters for the individual. It has been subsequently extended to account for prior uncertainty in model parameter values (22, 23).

The first attempt at employing the population Fisher information matrix (PFIM) in designing PPK studies was made by Wang and Endrenyi (5), who used NONMEM to obtain operational estimates of the PFIM for alternative designs. They took advantage of the theoretical property that both of the alternative variance–covariance matrices computed as part of the COVARIANCE step in NONMEM (the R^{-1} and S^{-1} matrices) converge asymptotically to the PFIM as sample size increases, given standard maximum likelihood estimation assumptions. They noted that, theoretically, the elements of the variance–covariance matrices should be inversely proportional to the number of subjects, and they confirmed this notion by evaluating the matrices for different sample sizes. They also noted that the sampling times determined by D-optimality are truly optimal for unbiased estimators, and that evaluation of the PFIM provides estimates of the precision but not of the accuracy of parameter estimates.

The importance of informative sampling for PPK parameter estimation was further investigated by Ette et al. (8) for a two-compartment model drug

administered by IV bolus. They proposed a method, the informative block random-ized design, that combined the efficiency of D-optimality criterion and the robust-ness afforded by random sampling. With this design, the concentration profile is divided into contiguous sampling blocks or intervals, and equal numbers of samples are chosen at random from each interval. The authors noted that clinical trial investigators would be more comfortable to sample within specified windows than to sample randomly without regard to any particular region of the plasma concen-tration–time profile. The informative block (profile) randomized (IBR) sampling approach also permits the use of mixed designs, in which fewer samples are obtained from some of the subjects in the study (11).

The theoretical basis for optimal sampling in population pharmacokinetics was advanced by Mentre et al. (24), who provided an analytical solution to the popula-tion Fisher information matrix (PFIM). This solution assumed that interindividual variability in parameters were independent (diagonal variance–covariance matrix), with a proportional residual error. Optimal designs were determined, for a given set of PPK parameter values. A further theoretical advance to account for uncer-tainty in the prior parameter values was provided by Tod et al. (13), who proposed several alternative cost functions based on the expected value of the determinant of the PFIM, to account for uncertainty in the prior values of the PPK parameters in determining optimal sampling times. This solution of the PFIM and its determinant was subsequently extended to account for heteroscedastic residual error models (with respect to mean parameters) by Retout and Mentre (25). The S-Plus and Matlab codes that solve the PFIM to determine optimal sampling times are now publicly available (26, 27).

One of the challenges in determining designs based on PFIM D-optimal criteria is the implementation of methods to identify the global minimum of the PFIM deter-minant. Several search algorithms for maximizing the determinant of the population Fisher information matrix (PFIM) have been proposed as a means of obtaining the most optimal PPK design. They include the Fedorov–Wynn algorithm, simulated annealing, and the Nelder–Mead simplex algorithm. Duffull et al. (28) noted that the response surface of the PFIM determinant with respect to parameter values could be highly convoluted, making it difficult to find a global minimum. Moreover, they noted that the PFIM (unlike the individual Fisher information matrix) is not invariant to model parameterization. They studied the ability of several alternative search algorithms to consistently identify the minimum and found that simulated annealing and a combination of nonadaptive random search and the simplex algo-rithm provided the best results.

A comparative evaluation of the efficacy of population designs determined using individual and population D-optimality criteria was investigated by Hooker et al. (29) for one- and two-compartment first-order absorption PK models, and a viral dynamics PD model. As would be expected, the sampling times determined by employing population D-optimality criteria were generally distributed around the individual D-optimal sampling times. These authors found that the accuracy and precision of model parameter population average and variance estimates were comparable for all the designs examined, under the assumption that residual error was known. They noted that the advantage of designs determined using population D-optimality is that they had fewer catastrophic estimates for some parameters, and that they permitted fewer samples per individual. These results suggest that

PPK designs with sampling times selected at random from time windows around times determined by individual D-optimality—similar to the informative block randomized design (8)—provide robust parameter estimates. A direct comparison of the efficiency of sampling designs determined using individual and population D-optimality is presented below.

12.4.2.1   Simulation Study Comparison of IBR and PFIM Designs

The IBR and PFIM D-optimal methods were compared using enoxaparin as an example (15). As discussed above, both designs are based on D-optimality. The example was based on PPK sampling designs developed for a clinical trial in which 30 mg of enoxaparin is administered as an intravenous bolus dose to 200 subjects, followed by five subcutaneous doses of 85 mg administered q12h (30). Designs that were optimal under constraints of maximum number of samples/subject were compared.

The population pharmacokinetics of enoxaparin was described by a one-compartment IV bolus model, the parameters of which are presented in Table 12.1. The interindividual variability parameters specify variances in the log-scale of the lognormally distributed PK parameters, and the residual error parameter specifies the variance of the proportional error. This model was implemented in ADAPT II using the Fortran code, provided in Appendix 12.1, which is identical to the 1COMPCL.FOR code provided as part of the software distribution, except for the residual error model.

The code in Appendix 12.1 was used to generate ADAPT II executable files for simulation and sample optimization. The interindividual variability parameters in Table 12.1 were ignored in determining the optimal sampling times. This is equivalent to determining the optimal sampling times for the typical individual, with PK parameters identical to the typical values of the fixed effect parameters in Table 12.1. The four optimal sampling times obtained from ADAPT II were at 0.5, 2.82, 50.35, and 60 hours, post-first dose. Two alternative IBR sampling designs (IBR-4A, and IBR-4B) were evaluated based on these sampling times. These sampling designs are presented in Table 12.2, along with the PFIM design proposed by Retout et al. (30).

TABLE 12.1   Enoxaparin Population Pharmacokinetic Parameters

Parameter	Units	Estimate
Fixed effect parameters		
CL.TV	L/h	0.708
V.TV	L	5.49
KA	h^{-1}	0.232
Random effect parameters (interindividual variability)		
CL.OM	—	0.175
V.OM	—	0.277
Residual error parameters		
ERR.VAR	—	0.0682

TABLE 12.2   Summary of Four Samples/Subject Designs Evaluated

Design	Description	Sampling Times[a] (h)
$\Xi_{\text{PFIM-4}}$	Population Fisher information matrix D-optimal design (4 samples/subject)	0.5, 4, 50.5, and 60
$\Xi_{\text{IBR-4A}}$	Informative block randomized design (4samples/subject); samples selected from discrete sets	{0.25, 0.5, 0.75, 1}, {2, 3, 4, 6, 8}, {50, 52, 54}, and {56, 58, 60}
$\Xi_{\text{IBR-4B}}$	Informative block randomized design (4samples/subject), samples selected from within continuous intervals	{0.25 − 1}, {2 − 8}, {50 − 54}, and {56 − 60}

[a] Sampling time is with respect to first dose. Curly braces denote set of times from which one sample is selected at random.

Design IBR-4A consists of four sets of discrete sampling time points, centered around the optimal sampling time points determined using ADAPT II. The sampling schedule for each subject is specified by selecting one sample at random from each of these sets. Likewise, design IBR-4B consists of four sampling time point intervals, and the sampling schedule for each subject is specified by a random draw from each uniform distribution specified by these intervals. The S-Plus and NONMEM code to compare the efficiency of sampling designs with four samples/subject are described below.

The S-Plus code in Appendix 12.2 can be used to automate the creation of template data sets that serve as input to NONMEM control files to simulate clinical trials. This code creates three template data sets: data set `SimData.PFIM_4.csv` is for the PFIM design described by Retout et al. (30), while `SimData.IBR_4A.csv` and `SimData.IBR_4B.csv` are for two alternative IBR designs.

The NONMEM control file in Appendix 12.3 provides a template for simulating PK data according to a given design that is specified by a template data set. In this control file, `SimData.Design.csv` is a placeholder for one of the above described template data sets. Multiple trials can be simulated by reproducing this file and changing the random seed (specified by the `zzzzz` place holder in Appendix 12.3) and the output table file name. A UNIX shell script that accomplishes this is given in Appendix 12.4. An alternative method of simulating multiple trials is using the "SUBPROBLEMS" option to the NONMEM `$SIMULATE` command. However, this latter option creates a single data file, requiring a single NONMEM control file to estimate the parameters for the multiple simulated clinical trials. The drawback of this approach is that the subsequent NONMEM run to estimate parameters may not terminate successfully if difficulty is encountered in estimating PK parameters for an intermediate trial.

In the approach adopted here, multiple data files are created, each containing the data for a single simulated clinical trial. Separate NONMEM estimation runs are performed for each data file, thus ensuring that the failure of a NONMEM run to terminate for a given simulated trial does not affect the ability to obtain estimates for the remaining trials.

The NONMEM control file for parameter estimation is provided in Appendix 12.5. The UNIX script provided in Appendix 12.6 creates a set of estimation files that are identical, except for the name of the input datafile. One set of NONMEM

estimation control files is created for each of the simulated designs being compared, so that a given set contains all the NONMEM estimation control files that correspond to the data simulated for a given sampling design. The parameter estimates resulting from a set of NONMEM estimation control files is extracted using a Perl script given in Appendix 12.7, which creates an ASCII file for each set of NONMEM estimation control files, with parameter estimates for each NONMEM run on a single line. Finally, the S-PLUS script in Appendix 12.8 is used to analyze the extracted NONMEM parameter estimates for all the designs being compared, and to plot the results of the analysis. Figures 12.1–12.4 present a graphical comparison of the accuracy and precision of NONMEM fixed and random effect parameter estimates from the alternative sampling designs. All designs performed similarly in the production of efficient PPK parameter estimates.
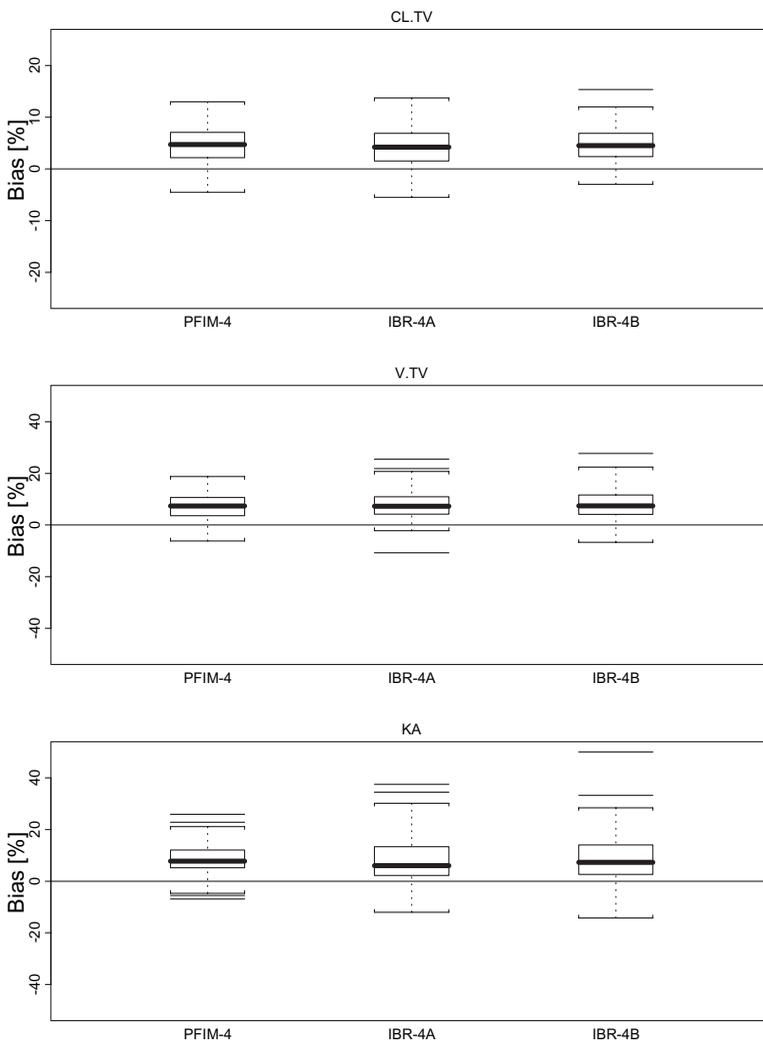
FIGURE 12.1 Comparison of bias in fixed effect parameter estimates for the PFIM-4, IBR-4A, and IBR-4B designs (see text for description of the designs).
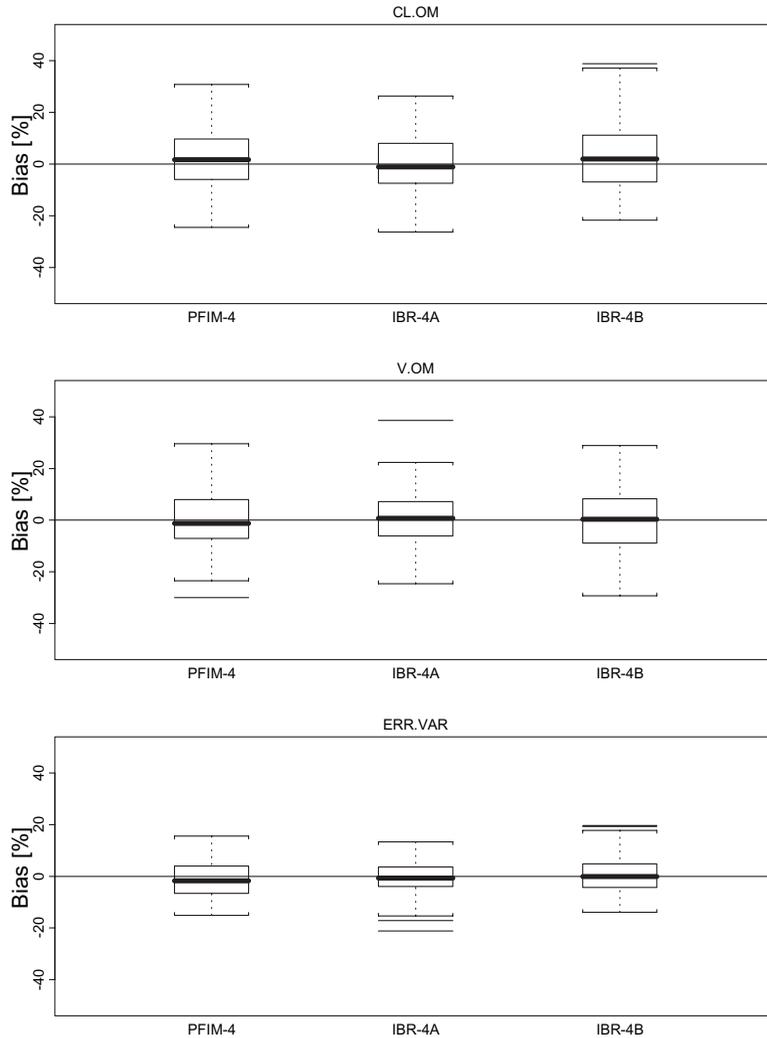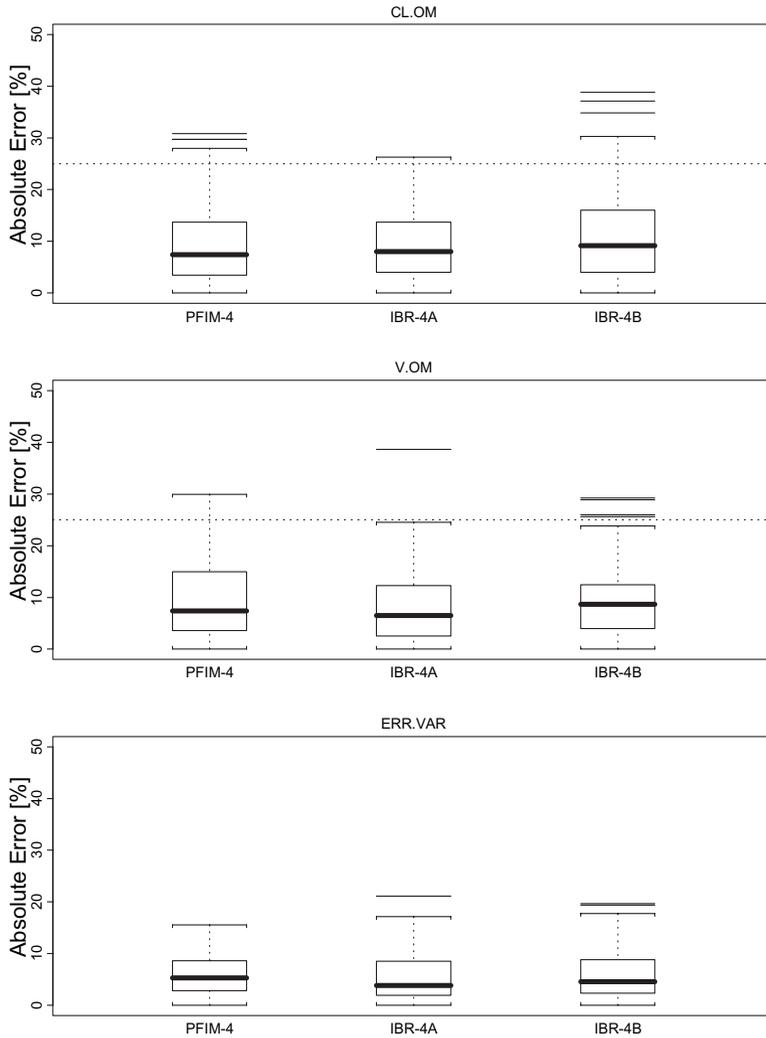
FIGURE 12.2 Comparison of bias in random effect parameter estimates for the PFIM-4, IBR-4A, and IBR-4B designs (see text for description of the designs).

12.4.2.2 Overall Assessment of the Performance of IBR and PFIM Designs

Designs for population studies should be pragmatic and should not overly interfere with the primary objectives of clinical trials (31). As a D-optimality criterion does not always provide pragmatic designs, heuristic designs based on D-optimal criterion should be evaluated using clinical trial simulation. Clinical trial simulation followed by parameter estimation is a better measure of optimality as it allows both the bias and precision to be evaluated, and these measures can be compared for each parameter. The results of our simulation study indicate that both IBR and PFIM D-optimal designs performed similarly in yielding efficient PPK parameter
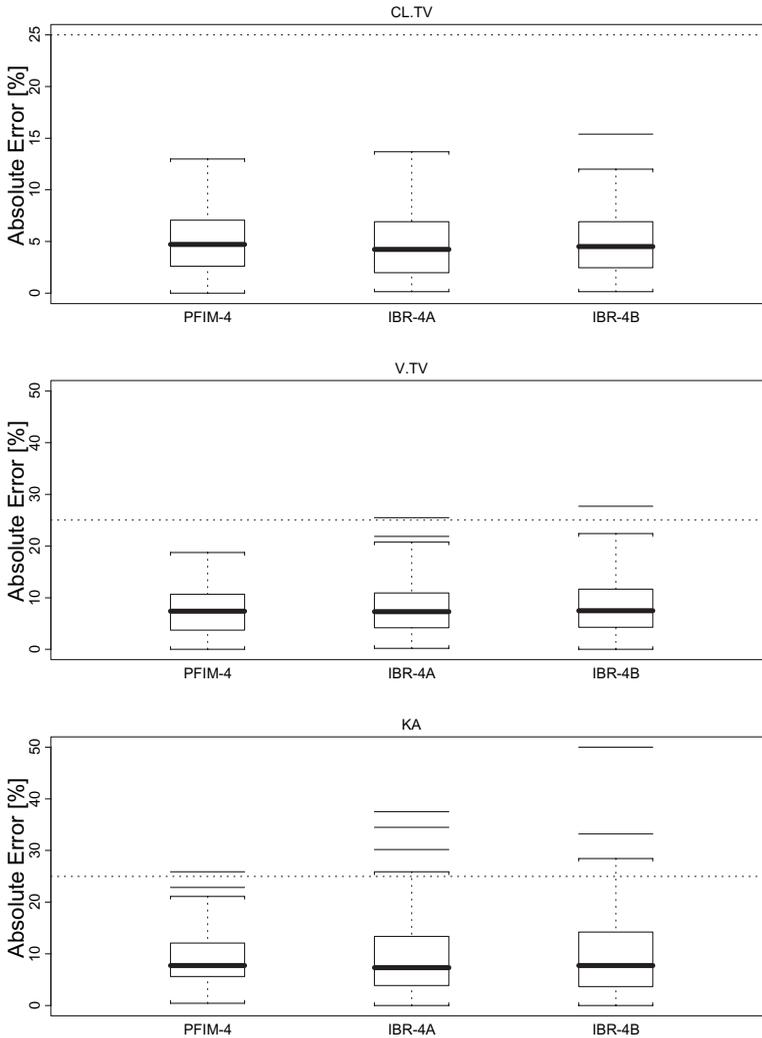
FIGURE 12.3   Comparison of precision in fixed effect parameter estimates for the PFIM-4, IBR-4A, and IBR-4B designs (see text for description of the designs).

estimates. The ease with which the IBR designs can be generated makes them preferable in drug development, where pragmatism and time are of great consideration. The IBR designs have been referred to as pragmatic designs (15). Pragmatic designs that achieve high efficiency in the estimation parameters should be used in the design of population PK studies. This takes on greater significance when mixed designs are implemented, as is usually the case, in later stage development clinical trials (15). The excellent performance of the IBR designs and the PFIM D-optimal designs in the estimation of PPK parameters points to the fact that enough samples were located in informative regions of the plasma concentration profile, enabling efficient parameter estimation.

FIGURE 12.4 Comparison of precision in random effect parameter estimates for the PFIM-4, IBR-4A, and IBR-4B designs (see text for description of the designs).

12.5 NUMBER OF SAMPLES PER SUBJECT

Most PPK studies are performed as adjuncts to clinical trials. There is the constant challenge to make the most from a PPK study performed within the context of the clinical trial without jeopardizing the primary objective of the trial. In practice, rarely is a balanced sampling design used in longitudinal PPK studies. Also, the number of individuals to be studied and the individual designs to be performed when there is a maximum cost (i.e., maximum number samples) to contend with may vary. Fadiran et al. (11) investigated the effect of various costs of PPK designs using mixed designs on the efficiency of PPK parameter estimation. A mixed design PPK study is one in which the number of samples/subject is not identical for all

subjects in a study. Parameter estimation efficiency was shown to deteriorate with designs having a lot of subjects contributing only one sample per subject. With such designs residual variability was underestimated with a corresponding positive bias in the estimation of volume terms of the two-compartment model drug. The authors cautioned against the artifact in parameter estimation by having an adequate number of subjects sampled more than once.

12.5.1   Sampling for Interoccasion Variability Estimation

The variance of an individual's PK observations about the individual-specific PK model on a given occasion (i.e., the intraindividual variability) can conceptually be factored into two components: variability of PK observations due to variability of PK parameters from occasion to occasion (interoccasion variability), and variability of PK observations about the individual PK model appropriate for the particular occasion (noise, PK model misspecification). To be sure, some interoccasion variability may be explained by interoccasion variation in individual time-varying covariates, but to the extent that it is not, it represents, along with the noise, the irreducible uncertainty in predicting, and hence controlling, drug concentrations. Drugs with narrow therapeutic indices and large interoccasion and intrasubject variability, for example, will be very difficult to control. If a PPK study consists of PK observations solely from individuals studied on only a single occasion, the interoccasion variability may appear incorrectly in the interindividual variability term and not in the intraindividual variability term. This may lead to inappropriate optimism about the ability to control individual therapy within the therapeutic range by using feedback (e.g., therapeutic drug monitoring, or simply adjusting dose according to observed drug effects), and also to a fruitless search for interindividual covariates that might explain the (spuriously inflated) interindividual variability. It is of the utmost importance to avoid this artifact by ensuring that at least a moderate subset of subjects contributing data to a PPK study contribute data from more than one occasion. Indeed, if this is done, one may hope to separately estimate the components of intraindividual variability (32).

12.6   SAMPLE SIZE AND STUDY POWER

Having the appropriate sample size is important for the efficient estimation of PPK parameters from a study. Ette and co-workers (10, 11, 16) have shown from simulation studies that robust estimates of clearance can be obtained with sample sizes ≥30, depending on the intersubject variability studied. Estimates of volume terms or volume-related terms such as intercompartmental clearance in a two-compartment model have been shown to require sample sizes ≥50 (10, 11, 16).

When a PPK study is designed to detect a difference between two subpopulations or to determine important covariates necessary to explain variability, attention should be paid to the sample size required for such a study. Simulation plays an important role in this situation, and Kowalski and Hutmacher (33) demonstrated the importance of using clinical trial simulations to assess the power to detect subpopulation differences in apparent drug clearance (CL/F) and sample size requirements for a PPK substudy (1) of a Phase 3 clinical trial. Two subpopulations were

investigated. The simulations were based on a PPK model developed from a Phase 1 healthy volunteer study. The key parameter of interest in the simulated study was CL/F, and a 40% reduction was considered to be of clinical significance. That is, this degree of reduction in CL/F would result in a need for dosage adjustment. It was also desired to detect this difference in as small as 5% of the total patient population, thus providing a framework for powering the study. Three hundred hypothetical clinical trials were simulated to determine the sample size necessary to detect 40% reduction in CL/F in a subpopulation of proportion $p = 0.05$ or $p = 0.10$ with at least 90% power. Sample sizes of 150 and 225 were investigated. The power of the study was estimated as the percentage of trials out of 300 in which statistically significant ($\alpha = 0.05$) difference in CL/F was observed using the likelihood ratio test.

To obtain the empirical estimates of α, Kowalski and Hutmacher (33) simulated 300 clinical trials for each combination of sample size and p, where the proportional reduction in CL/F (ϕ) was fixed to zero. Covariate and base models were fitted to each of the trials and the likelihood χ^2 ratio tests were performed at the 5% level of significance. The percentage of trials where a statistically significant difference in CL/F was observed provided an empirical estimate of α (i.e, H_0: $\phi = 0$ is rejected when H_0 is true). The data were analyzed with the NONMEM population pharmacokinetics/pharmacodynamics analysis software. The results suggested that an approximate nine-point change in the objective function should be used to assess statistical significance at the 5% level rather than the commonly used χ^2 critical value of 3.84 for one degree of freedom.

Thus, the importance of determining power in PPK studies for detecting an important covariate in a PPK study was highlighted, and attention was drawn to the use of the likelihood ratio test in PPK model development.

12.6.1 Usage of Likelihood Ratio Test

Ette et al. (34) in their review of the work of Kowalski and Hutmacher (33) drew attention to the risk inherent in the use of the likelihood ratio test (LRT) in the analysis of simulated trials, particularly in the context of nonlinear mixed effects modeling. The authors stated:

> If minus twice the log likelihood associated with the fit of a full model, A, with $q + r$ parameters is designated ℓ_A, and a reduced version of this model (model B) with p parameters has minus twice the log likelihood ℓ_B, the difference in minus twice the log likelihoods, ($\ell_A - \ell_B$) is asymptotically χ^2 distributed with q degrees of freedom. This formulation is widely used to assess the statistical significance level of the parameters associated with the q degrees of freedom.

> For the determination of the significance level of fixed effects, the LRT is known to be anti-conservative, i.e., the empirical p value will be greater than the nominal p value [35]. Generally, as the number of parameters (degrees of freedom) being tested increases, the more anti-conservative the test.

> Conversely, Stram and Lee [36] noted that the LRT tends to be asymptotically conservative for the assessment of random effects significance level. In this context, the conservative nature is attributable to the null hypothesis consisting of setting

the random effect at a boundary condition, i.e., zero. While the inaccuracy in p value is modest when the number of random effects being tested is small, the conservativeness increases with an increase in the number of random effects being tested.

Wählby et al. [37] explored via simulation a number of factors influencing the disparity between nominal and actual significance level of tests for covariate (fixed) effects in nonlinear mixed effects models using the software NONMEM. Approximation method (first order [FO] versus first order conditional estimation [FOCE] with $\eta–\varepsilon$ interaction), sampling frequency, and magnitude and nature of residual error were determined to be very influential on the bias associated with the p value. An important finding was that the use of the FOCE method with $\eta–\varepsilon$ interaction resulted in reasonably close agreement of actual and nominal significance levels, whereas the application of the LRT after estimation using the FO approximation generally resulted in marked bias in p values.

The implications of the disparity between empirical and nominal significance levels of the likelihood ratio test in mixed effects modeling and simulation are clear; however, definitive solutions or corrections are not. While the significance of random effects is not generally the subject of interest in a simulation, the bias in likelihood ratio test–determined p value for fixed effects could be very influential on trial simulation findings. Thus, simulation exercises should provide for determination of empirical p values to avoid faulty conclusions about power and sample size.

12.6.2   Determining the Power of PPK Study for Characterizing Drug–Drug Interaction

Pharmacokinetic drug–drug interactions (DDIs) occur when the pharmacokinetics of a drug is affected by the concomitant administration of another drug. Characterization of DDIs for an investigative drug therefore involves assessing the potential for coadministered drugs to affect the pharmacokinetics of the investigative drug, as well as an assessment of the potential of the investigative drug to affect the pharmacokinetics of coadministered drugs.

It is clearly not practical to assess DDIs for all possible coadministered drugs by separate studies. The DDI potential for an investigative drug is conventionally assessed with well-controlled healthy volunteer studies of the investigative drug with drugs that are relatively specific prototypic inhibitors, inducers, or substrates of a given cytochrome P450 isozyme (38). DDI studies are generally designed to characterize the worst case scenario, so that if the potential for a DDI exists, it will be detected by the study. Therefore, it is possible that the DDI with a drug that is expected to be coadministered with the investigative drug is not as severe as the DDI in the well-controlled study with the prototypic drug. Moreover, as most DDI studies focus on interactions involving Phase 1 drug–metabolizing enzymes, other interactions involving Phase 2 drug–metabolism may remain undetected, as may interactions involving drug transporters.

It is therefore desirable to augment the information gained from well-controlled DDI studies with DDI information in the actual target patient population. Characterization of DDI in the target patient population using PPK analyses is in fact supported by the FDA's guidance on clinical DDI studies (39). The advantage of investigating DDI using population pharmacokinetics is that clinically significant

interactions are characterized, rather than the worst case interactions, while the main disadvantage is that detailed information on coadministered drugs (dose, frequency, and time of administration) is generally not available. Another disadvantage noted by the FDA guidance on DDI is that the ability to detect DDI using population pharmacokinetics has not been adequately studied. The investigation described next attempts to address this last deficiency.

12.6.2.1 Study Objective and Design

The objective of the analysis was to investigate the effect of PPK design on the power to detect a DDI affecting the pharmacokinetics of the investigational drug.

It was assumed that the pharmacokinetics of the investigational drug was described by a one-compartment model with first-order absorption, the parameter values of which are given in Table 12.3. Study design and drug attributes investigated were sample size (50, 100, 150, and 200 subjects), percentage of subjects on concomitant interacting drug (10%, 30%, 50%, 70%, and 90%), and interindividual variability (IIV) in PK clearance and apparent volume of distribution (30%, 50%, and 70%). It was assumed that concomitant administration of the interacting drug would decrease the clearance by 30%.

12.6.2.2 Simulation and Data Analysis

The PPK model was implemented in NONMEM, and data sets simulated with the model were then fitted using both a full (true) PPK model that included the DDI effect of the concomitantly administered interacting drug, and a reduced (null) model that did not contain the interaction effect. The DDI effect of the interacting drug was incorporated into the PK model as a covariate on clearance. The power of a given design to detect the presence of interaction was determined from using the likelihood ratio test (LRT) (See Section 12.6.1). As the full and reduced models differ by only a single parameter (the covariate effect of interacting drug on clearance), the reduced model is nested within the full model and has 1 degree of freedom less than the full model. Therefore, the DDI should be considered to be statistically significant at the 5% level, provided that the NONMEM objective function value (OFV) for the full model is at least 3.84 points less than that of the reduced model (as the NONMEM objective function value is equivalent to $-2\log$-likelihood).

However, as noted in the discussion on the LRT (Section 12.6.1), the test tends to be conservative for fixed effects, suggesting that the actual critical value for the LRT statistic may be larger than 3.84. Moreover, the number of samples per subject and the sample size may also affect the theoretical critical value, as the likelihood ratio is asymptotically χ^2 distributed. Previous work also indicated that the likelihood

TABLE 12.3 Population Pharmacokinetic Parameters of Investigational Drug

Parameter	Geometric Average	Interindividual Variability (%CV)
Clearance (CL), L/h	1.73	30, 50, 70
Volume (V), L	30	30, 50, 70
Absorption rate (K_a), h^{-1}	1	—

ratio was also sensitive to the linearization method used to obtain the NONMEM OFV (37, 40).

The Type I error (rejection of the reduced model in favor of the full model) that would result from the use of the theoretical critical value was assessed for each of the designs considered, and for three alternative NONMEM linearization methods: first-order (FO), first-order conditional estimation (FOCE), and first-order conditional estimation with interaction (FOCEI). Type I error rates were assessed by empirical determination of the probability of rejection of the reduced model, given that the reduced model was the correct model. Data sets were simulated with the reduced model (FO, 1000 data sets; FOCE/FOCEI, 200 data sets) and fitted using the full and reduced models. The empirical Type I error was determined as the percentage of simulated data sets for which a LRT statistic of 3.84 or greater was obtained. The 3.84 critical value for the LRT statistic corresponds to a significance level of 5%, for a χ^2 distribution with 1 degree of freedom (for the one extra parameter in the full model). The LRT statistic was calculated as the difference between the NONMEM objective function values of the reduced and full models. The results of these simulations were also used to determine an empirical critical value that would result in the Type I error rate equal to the nominal 5% value.

The power of a design with a given set of attributes (sample size, percentage of subjects on interacting drug, and percentage IIV) was assessed by determining the probability of rejecting the reduced model, given that the full model was the correct model. In this case, data sets were simulated with the full model and fitted using the full and reduced models. The theoretical power was determined as the percentage of data sets for which a LRT statistic of 3.84 was obtained. The empirical power was determined by using the empirical critical value, instead of the theoretical value of 3.84.

12.6.2.3 Study Outcome

The results of the Type I error determination are summarized in Figure 12.5. Type I error rates were calculated provided that the LRT could be calculated for at least 100 of the simulated data sets. This criterion was met for all designs with the FO and FOCE methods, but the Type I error is not reported for several designs with the FOCEI method. The LRT statistic could not be determined for a larger number of runs with the FOCEI method because achieving convergence with this method is more difficult than with the other methods. In this analysis, no attempt was made to improve the convergence rate by manually adjusting the initial parameter estimates. Consistent with the previously reported findings (37), the Type I error was greatest for the FO method and least for the FOCEI method. The Type I error rate also increased with increase in IIV, and differences between the FOCEI method and the other two methods were more pronounced at higher IIV. Notably, the empirical Type I error for the FOCEI method was close to the nominal value of 5%, for all designs in which it was estimated.

In the present study, the power of PPK study designs to detect DDI was estimated using the empirical critical values, given the inflated Type I error obtained using the theoretical critical value. An empirical critical value was determined for each combination of study design, drug IIV, and NONMEM estimation method (Table 12.4). The empirical critical value for a given combination was set at the 95th percentile value of the LRT statistic distribution.
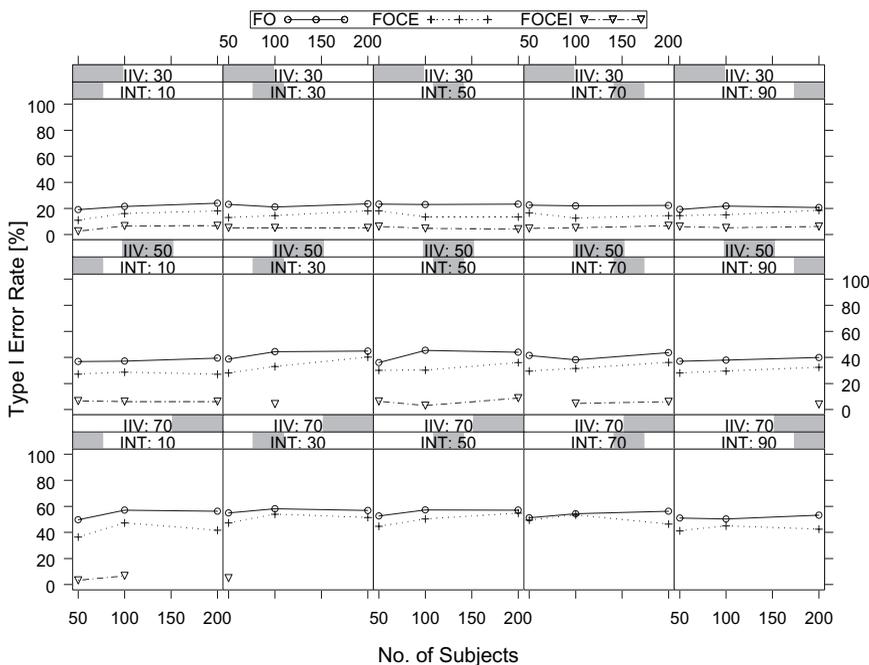
FIGURE 12.5 Type I error for the NONMEM estimation methods (FO, FOCE, and FOCEI) for each design. Each panel represents a level of interindividual variability (IIV, constant along a row), and percentage of subjects on interacting drug (INT, constant along a column). The effect of sample size is shown within each panel.

A comparison of the power to determine DDI with theoretical and empirical critical values for the FO method is presented in Figure 12.6. As expected, the theoretical power to determine DDI is higher than the empirically determined power, given the inflated Type I error with the theoretical critical value. The empirical power is a more accurate representation of the true power.

The effect of study design, IIV in PK parameters, and estimation method on power to detect DDI is presented in Figure 12.7. Much of the differences between the estimation methods seen with the theoretical critical value appear to have been eliminated, particularly differences between the FO and FOCE methods. The empirical power could not be determined with the FOCEI method due to failure of a sufficient number of estimation runs to converge. However, for cases in which the power with the FOCEI method could be determined, it was consistently greater than the power obtained with the other two methods.

The power to detect DDI is most profoundly affected by the IIV in PK parameters. A change of 40% in the IIV (30% to 70%) had a greater impact on the power than a fourfold change in sample size (50 to 200 subjects), or a ninefold change in INT, the percentage of subjects on the interacting drug (10% to 90%). The next most important factor was sample size, followed by INT. With respect to INT, the power was greatest when 50% of subjects were on the interacting drug.
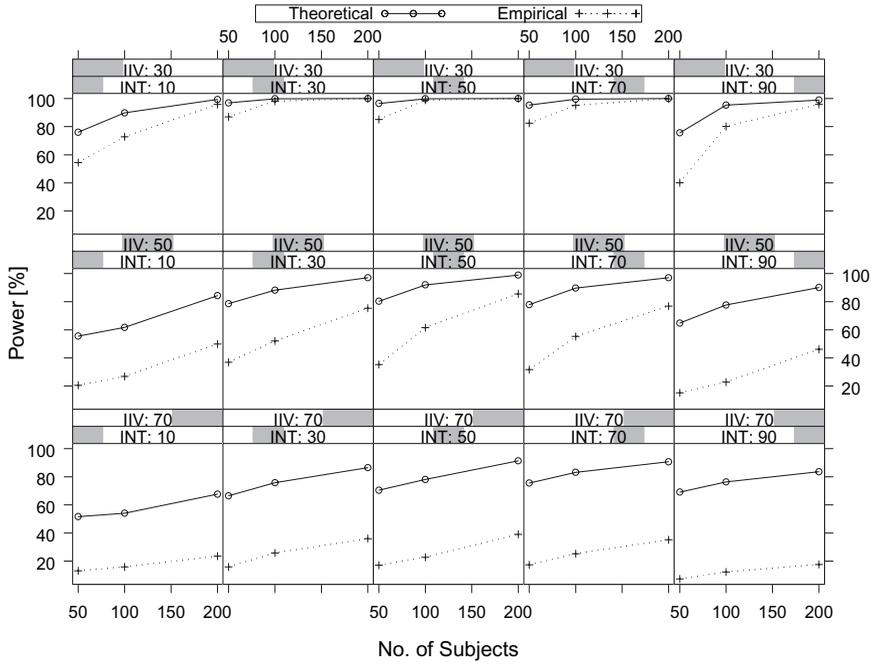
FIGURE 12.6 Theoretical and empirical power to determine DDI with the FO estimation method. Each panel represents a level of interindividual variability (IIV, constant along a row), and percentage of subjects on interacting drug (INT, constant along a column). The effect of sample size is shown within each panel.
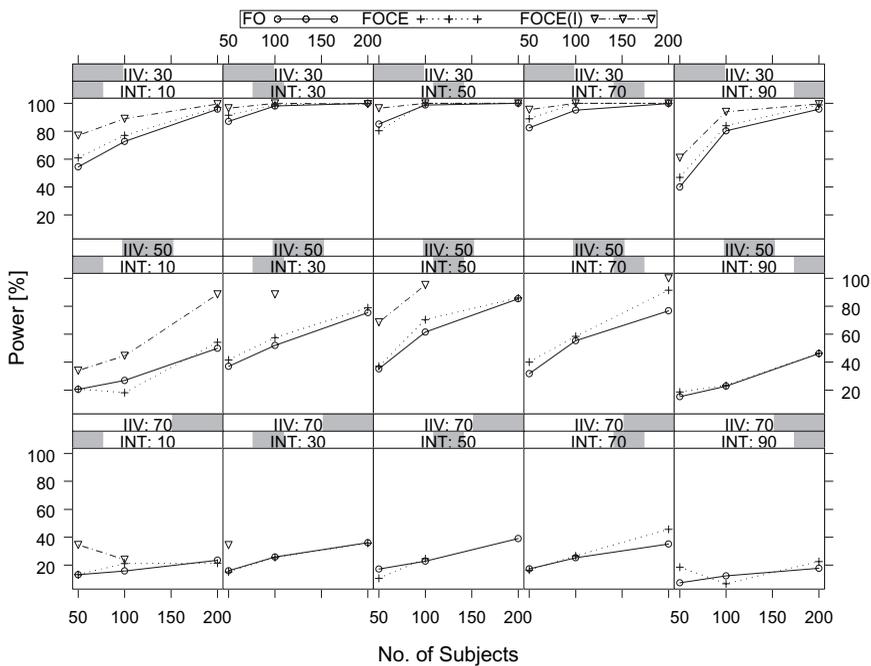
FIGURE 12.7 Empirically determined power to determine DDI with the FO, FOCE, and FOCEI estimation method. Each panel represents a level of interindividual variability (IIV, constant along a row), and percentage of subjects on interacting drug (INT, constant along a column). The effect of sample size is shown within each panel.

TABLE 12.4 Population Pharmacokinetic Parameters of Investigational Drug

Sample Size	Subjects on Interacting Drug (%)	Interindividual Variability (%CV)	FO Method		FOCE Method		FOCEI Method	
			Type I Error	Empirical Critical Value	Type I Error	Empirical Critical Value	Type I Error	Empirical Critical Value
50	10	30	19.0	8.8	11.0	5.7	2.5	3.0
50	10	50	36.9	16.4	27.1	11.2	6.6	4.0
50	10	70	49.7	25.4	36.5	16.5	3.1	2.9
50	30	30	23.1	8.7	13.0	6.7	5.0	3.7
50	30	50	38.8	19.2	28.0	12.9	4.4	3.6
50	30	70	55.0	36.0	47.2	28.0	4.9	3.9
50	50	30	23.3	9.9	18.0	8.9	6.0	4.5
50	50	50	36.0	20.7	30.0	14.7	6.2	4.1
50	50	70	52.6	34.1	44.7	29.7	11.3	5.5
50	70	30	22.5	10.2	16.5	6.7	4.5	3.8
50	70	50	41.4	19.9	29.5	12.8	5.4	3.5
50	70	70	51.2	33.9	49.5	28.9	5.0	3.4
50	90	30	19.2	9.8	14.5	7.4	6.1	4.1
50	90	50	37.0	15.8	28.0	12.1	2.7	2.9
50	90	70	51.1	24.2	41.2	16.4	9.8	4.3
100	10	30	21.6	9.4	16.0	7.3	6.5	4.4
100	10	50	37.2	17.8	28.6	13.3	6.1	4.2
100	10	70	57.1	29.1	47.2	19.3	6.6	4.4
100	30	30	21.1	9.3	14.5	7.5	5.0	3.8
100	30	50	44.4	21.6	33.0	13.4	4.1	3.7
100	30	70	58.1	41.1	53.9	34.5	4.3	3.8
100	50	30	22.9	9.9	13.5	7.4	4.5	3.7

100	50	50	45.4	21.4	30.2	13.9	3.0	3.2
100	50	70	57.3	44.8	50.3	34.2	8.3	6.1
100	70	30	21.9	10.1	12.6	6.8	5.0	3.5
100	70	50	38.1	19.3	31.5	15.1	4.6	3.9
100	70	70	54.4	37.9	53.3	31.6	1.3	2.7
100	90	30	21.9	8.5	15.0	7.8	5.1	3.6
100	90	50	37.9	19.9	29.5	16.7	8.9	5.9
100	90	70	50.3	31.9	44.9	34.4	7.7	3.7
200	10	30	24.0	10.4	18.0	7.4	6.6	4.6
200	10	50	39.5	19.9	27.0	13.1	6.1	3.6
200	10	70	56.3	34.5	41.6	23.9	6.9	6.9
200	30	30	23.5	9.9	18.0	8.1	5.0	3.8
200	30	50	44.8	26.7	40.2	18.1	0.0	0.7
200	30	70	56.9	49.4	51.5	41.5	15.4	4.7
200	50	30	23.3	9.6	13.5	7.1	4.0	3.6
200	50	50	44.0	23.4	36.0	15.7	8.7	4.9
200	50	70	57.1	54.5	54.9	41.9	2.9	3.2
200	70	30	22.2	9.4	14.5	8.2	6.6	4.2
200	70	50	43.6	25.9	36.2	14.5	6.0	4.2
200	70	70	56.3	47.4	46.4	29.3	0.0	1.5
200	90	30	20.7	9.2	18.5	6.3	6.1	4.1
200	90	50	39.9	21.3	32.3	16.5	3.9	3.4
200	90	70	53.3	39.2	42.6	25.6	3.6	3.1

12.6.2.4 Summary of Study Findings

The above results suggest that it is reasonable to expect that DDI would be detected using the population approach for drugs with low or moderate IIV, but very large sample sizes would be needed for high variability drugs. It is recommended that simulation study such as the one presented here be performed to support a claim of a lack of DDI.

12.7 STUDY EXECUTION AND IMPACT ON PARAMETER ESTIMATION EFFICIENCY

Sun et al. (41), using the informative block (profile) randomized design, investigated the effect of sample time recording errors (both systematic and random) on the estimation of PPK parameters for a drug exhibiting two-compartment pharmacokinetics, for both single and multiple administrations. The PK profile was divided into three blocks and each subject was sampled across the blocks, providing two samples per block. They observed that negative systematic error in the recording of sample times resulted in efficient estimation of volume terms, while positive systematic error favored the efficient estimation of the clearance terms. These errors resulted in sufficient samples being located in critical regions for the estimation of volume terms (negative systematic error) and clearance terms (positive systematic error). Overall, they found that the efficiency in the estimation of clearance was not severely compromised for moderate sampling time recording errors.

12.8 SUMMARY

The quality of the PPK parameter estimates is a function of experimental design, and design features and other relevant issues that should be addressed in designing PPK studies for efficient parameter estimation are discussed. Therefore, effective use of the population approach to characterize population pharmacokinetics demands that consideration should be given to how studies can be designed to obtain efficient estimates of PPK parameters of a drug studied. In addition, we show that DDI can be characterized using the population approach for drugs with low or moderate IIV, but very large sample sizes are needed for high variability drugs.

REFERENCES

1. US Food and Drug Administration, *Guidance for Industry: Population Pharmacokinetics*. US Department of Health and Human Services, February 1999.
2. Y. Hashimoto and L. B. Sheiner, Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis. *J Pharmacokinet Biopharm* **19**:333–353 (1991).
3. M. K. Al-Banna, A. W. Kelman, and B. Whiting, Experimental design and efficient parameter estimation in population pharmacokinetics. *J Pharmacokinet Biopharm* **18**:347–360 (1990).

4. E. I. Ette, *Efficient Parameter Estimation in Preclinical Animal Pharmacokinetics*. Ph.D. thesis. University of Glasgow, Glasgow, Scotland, 1991.

5. J. Wang and L. Endrenyi, A computationally efficient approach for the design of population pharmacokinetic studies. *J Pharmacokinet Biopharm* **20**:279–294 (1992).

6. E. I. Ette, A. W. Kelman, C. A. Howie, and B.Whiting, Interpretation of simulation studies for efficient estimation of population pharmacokinetic parameters. *Ann Pharmacother* **27**:1034–1039 and Correction **27**:1548 (1993).

7. E. I. Ette, A. W. Kelman, C. A. Howie, and B. Whiting, Inter-animal variability and parameter estimation in preclinical animal pharmacokinetic studies. *Clin Research Regul Affairs* **11**:121–139 (1994).

8. E. I. Ette, H. Sun, and T. M. Ludden, Design of population pharmacokinetic studies. *Proc Am Stat Assoc* (*Biopharmaceut Sec*): 487–489 (1994).

9. E. I. Ette, A. W. Kelman, C. A. Howie, and B. Whiting, Efficient experimental design and estimation of population pharmacokinetic parameters in preclinical animal studies. *Pharm Res* **12**:729–737 (1995).

10. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional pharmacokinetic studies: implications for pediatric and animal studies. *Clin Res Regul Affairs* **13**:133–165 (1996).

11. E. O. Fadiran, C. D. Jones, and E. I. Ette, Designing population pharmacokinetic studies: performance of mixed designs. *Eur J Drug Metab Pharmacokinet* **25**:231–239 (2000).

12. N. E. Johnson, J. R. Wade, and M. O. Karlsson, Comparison of some practical sampling strategies for population pharmacokinetic studies. *J Pharmacokinet Biopharm* **24**:245–272 (1996).

13. M. Tod, F. Mentre, Y. Merle, and A. Mallet, Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. *J Pharmacokinet Biopharm* **26**:689–716 (1998).

14. S. Retout, S. Duffull, and F. Mentre, Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comput Methods Programs Biomed* **65**:141–151 (2001).

15. A. Roy and E. I. Ette, A pragmatic approach to the design of population pharmacokinetic studies. *Am Assoc Pharmceut Sci J* **7**:E408–E420 (2005).

16. E. I. Ette, H. Sun, and T. M. Ludden, Balanced designs in longitudinal population pharmacokinetic studies. *J Clin Pharmacol* **38**:417–423 (1998).

17. D. Z. D'Argenio, Optimal sampling times for pharmacokinetic experiments. *J Pharmacokinet Biopharm* **9**:739–756 (1981).

18. L. Endrenyi, Design of experiments for estimating enzyme and pharmacokinetic experiments, in *Kinetic Data Analysis of Enzyme and Pharmacokinetic Experiments*, L. Endrenyi (Ed.). Plenum Press, New York, 1981, pp. 137–167.

19. F. Mori and J. J. Di Stephano III, Optimal nonuniform sampling interval and test-input design for identification of physiological systems from very limited data. *IEEE Trans Automatic Control* **AC-24**(6):893–900 (1979).

20. E. M. Landaw, Optimal design for individual parameter estimation in pharmacokinetics, in *Variability in Drug Therapy: Description, Estimation, and Control*, M. Rowland, L. Sheiner, and J. L. Steimer (Eds.). Raven Press, New York, 1985, pp. 187–200.

21. D. Z. D'Argenio and A. Shumitzky, *ADAPT II User's Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software*. Biomedical Simulations Resource, Los Angeles, 1997.

22. D. Z. D'Argenio, Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Math Biosci* **99**:105–118 (1990).

23. M. Tod, C. Padoin, K. Louchahi, B. Moreau-Tod, O. Petitjean, and G. Perrit, Implementation of OSPOP, an algorithm for the estimation of optimal sampling times by the ED, EID, and API criteria. *Comput Methods Programs Biomed* **50**:13–22 (1996).

24. F. Mentre, A. Mallet, and D. Baccar, Optimal design in random-effects regression models. *Biometrika* **84**:429–442 (1997).

25. S. Retout and F. Mentre, Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *J Biopharm Stat* **13**:209–227 (2003).

26. F. Mentre and S. Retout, *S-PLUS Code for PFIMOPT*. `http:/www.bichat.inserm.fr/equipes/Emi0357/download.html`.

27. S. Duffull, *MATLAB Code for POPT*. `http://www.uq.edu.au/pharmacy/sduffull/POPT.html`.

28. S. B. Duffull, S. Retout, and F. Mentre, The use of simulated annealing for finding optimal population designs. *Comput Methods Programs Biomed* **69**:25–53 (2001).

29. A. C. Hooker, M. Foracchia, M. G. Dobbs, and P. Vivini, An evaluation of population D-optimal designs via pharmacokinetic simulations. *Ann Biomed Eng* **31**:99–111 (2003).

30. S. Retout, F. Mentre, and R. Bruno, Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Stat Med* **21**:2673–2679 (2002).

31. L. Aarons, L. P. Balant, F. Mentre, P. L. Morselli, M. Rowland, J.-L. Steimer, and S. Vozeh, Practical experience and issues in designing and performing population pharmacokinetic/pharmacodynamic studies. *Eur J Clin Pharmacol* **49**:251–254 (1996).

32. M. O. Karlsson and L. B. Sheiner, The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Biopharm* **21**:735–750 (1993).

33. K. G. Kowalski and M. M. Hutmacher, Design evaluation for a population pharmacokinetic study using clinical trial simulations: a case study. *Stat Med* **20**:75–91 (2001).

34. E. I. Ette, C. J. Godfrey, S. Ogenstad, and P. J. Williams, Analysis of simulated trials, in *Clinical Trial Simulations*, H. C. Kimko and S. Duffil (Eds.). Marcel Dekker, New York, 2002, pp. 131–151.

35. J. C. Pinheiro and D. M. Bates (Eds.), *Mixed-Effects Models in S and S-PLUS*. Springer, New York, 2000.

36. D. O. Stram and J. W. Lee, Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**:1171–1177 (1994).

37. U. Wählby, E. N. Jonsson, and M. O. Karlsson, Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn* **28**:231–252 (2001).

38. T. D. Bjornsson, J. T. Callaghan, H. J. Einolf, et al., The conduct of in vitro and in vivo drug–drug interaction studies: a PhRMA perspective. *J Clin Pharmacol* **43**:443–469 (2003).

39. *FDA Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling.* US Department of Health and Human Services, November 1999.

40. U. Wahlby, M. R. Bouw, E. N. Jonsson, and M. O. Karlsson, Assessment of type I error rates for the statistical sub-model in NONMEM. *J Pharmacokinet Pharmacodyn* **29**:251–269 (2002).

41. H. Sun, E. I. Ette, and T. M. Ludden, On the recording of sample times and parameter estimation from repeated measures pharmacokinetic data. *J Pharmacokinet Biopharm* **24**:245–263 (1996).

APPENDIX 12.1   ADAPT II FORTRAN CODE SPECIFYING A 1-COMPARTMENT 1ST-ORDER ABSORPTION MODEL

Proportional Error Given by Coefficient of Variation

```
C****************************************************************
C
C                   MODEL: 1COMPCL_PropErr
C     Modified Version of MODEL: 1COMPCL in ADAPT II (Release 4)
C     Residual error model changed to Proportional error...
C      ...specified in terms of CV
C****************************************************************
C                                                              *
C     This file contains the Fortran subroutines listed below in   *
C     which the user must enter the relevant equations and constants.*
C     Consult the user's manual for details concerning the format for *
C     entered equations and definition of symbols.                *
C                                                              *
C       1. DiffEq- System differential equations.             *
C       2. Amat  - System state matrix.                       *
C       3. Output- System output equations.                   *
C       4. Symbol- Parameter symbols and model constants.     *
C       5. Varmod- Error variance model equations.            *
C       6. Prior - Parameter mean and covariance values       *
C       7. Sparam- Secondary parameters                       *
C                                                              *
C****************************************************************

      Subroutine DIFFEQ(T,X,XP)
      Implicit None

      Include 'globals.inc'
      Include 'model.inc'

      Real*8  T,X(MaxNDE),XP(MaxNDE)

CC
C----------------------------------------------------------------C
C 1. Enter Differential Equations Below {e.g. XP(1) = -P(1)*X(1) } C
C----c-----------------------------------------------------------C

C     XP(1) = -(P(1)/P(2))*X(1) + P(3)*X(2) + R(1)
C     XP(2) = -  P(3)*X(2)

C----------------------------------------------------------------C
C----------------------------------------------------------------C
C
      Return
      End
```

```
C###############################################################C

      Subroutine AMAT(A)
      Implicit None

      Include 'globals.inc'
      Include 'model.inc'

      Integer I,J
      Real*8 A(MaxNDE,MaxNDE)

      DO I=1,Ndeqs
         Do J=1,Ndeqs
            A(I,J) = 0.0D0
         End Do
      End Do

CC
C----------------------------------------------------------------C
C 2. Enter non zero elements of state matrix {e.g. A(1,1) = -P(1) } C
C----c-----------------------------------------------------------C


C----------------------------------------------------------------C
C----------------------------------------------------------------C
C
      Return
      End


C###############################################################C

      Subroutine OUTPUT(Y,T,X)
      Implicit None

      Include 'globals.inc'
      Include 'model.inc'

      Real*8 Y(MaxNOE),T,X(MaxNDE)

CC
C----------------------------------------------------------------C
C 3. Enter Output Equations Below {e.g. Y(1) = X(1)/P(2) }      C
C----c-----------------------------------------------------------C

C Note: X(1) and X(2) are the amounts in the central and absorption
C   compartments, respectively.

      Y(1)=X(1)/P(2)
```

```
C--------------------------------------------------------------------C
C--------------------------------------------------------------------C
C
      Return
      End

C############################################################C

      Subroutine SYMBOL
      Implicit None

      Include 'globals.inc'
      Include 'model.inc'

CC
C--------------------------------------------------------------------C
C 4.                  Enter as Indicated                            C
C----c---------------------------------------------------------------C

     NDEqs   = 2 ! Enter # of Diff. Eqs.
     NSParam = 3 ! Enter # of System Parameters.
     NVparam = 1 ! Enter # of Variance Model Parameters.
     NSecPar = 3 ! Enter # of Secondary Parameters.
     NSecOut = 0 ! Enter # of Secondary Outputs (not used).
     Ieqsol  = 94 ! Indicates a built-in compartment model.
     Descr   = '1CmptAbs_CL - proportional variance'

C--------------------------------------------------------------------C
C--------------------------------------------------------------------C
C
CC
C--------------------------------------------------------------------C
C 4. Enter Symbol for Each System Parameter (eg. Psym(1)='Kel')     C
C----c---------------------------------------------------------------C

      PSym(1) = 'CLt'
      PSym(2) = 'Vc'
      PSym(3) = 'Ka'

C--------------------------------------------------------------------C
C--------------------------------------------------------------------C
CC
C--------------------------------------------------------------------C
C 4. Enter Symbol for Each Variance Parameter {eg: PVsym(1) = 'Sigma'} C
C----c---------------------------------------------------------------C

      PVsym(1) = 'CV'

C--------------------------------------------------------------------C
```

```
C----------------------------------------------------------------------C
CC
C----------------------------------------------------------------------C
C 4. Enter Symbol for Each Secondary Parameter {eg: PSsym(1) = 'CLt'} C
C----c-----------------------------------------------------------------C

        PSsym(1) = 'Kel'
        PSsym(2) = 'LAM1'
        PSsym(3) = 't1/2-LAM1'


C----------------------------------------------------------------------C
C----------------------------------------------------------------------C
C
        Return
        End


C#############################################################C

        Subroutine VARMOD(V,T,X,Y)
        Implicit None

        Include 'globals.inc'
        Include 'model.inc'

        Real*8 V(MaxNOE),T,X(MaxNDE),Y(MaxNOE)

CC
C----------------------------------------------------------------------C
C 5. Enter Variance Model Equations Below                              C
C       {e.g. V(1) = (PV(1) + PV(2)*Y(1))**2 }                         C
C----c-----------------------------------------------------------------C

        V(1) = (PV(1)/100*Y(1))**2

C----------------------------------------------------------------------C
C----------------------------------------------------------------------C
C
        Return
        End


C#############################################################C

        Subroutine PRIOR(Pmean,Pcov,ICmean,ICcov)
        Implicit None

        Include 'globals.inc'
        Include 'model.inc'
```

```fortran
      Integer I,J
      Real*8 Pmean(MaxNSP+MaxNDE), ICmean(MaxNDE)
      Real*8 Pcov(MaxNSP+MaxNDE,MaxNSP+MaxNDE),
      ICcov(MaxNDE,MaxNDE)

CC
C------------------------------------------------------------------C
C 6. Enter Nonzero Elements of Prior Mean Vector                   C
C        { e.g. Pmean(2) = 10.0 }                                  C
C----c-------------------------------------------------------------C


C------------------------------------------------------------------C
C------------------------------------------------------------------C
CC
C------------------------------------------------------------------C
C 6. Enter Nonzero Elements of Covariance Matrix (Lower Triang.)  C
C        { e.g. Pcov(2,1) = 0.25 }                                 C
C----c-------------------------------------------------------------C


C------------------------------------------------------------------C
C------------------------------------------------------------------C
C
      Return
      End


C###############################################################C

      Subroutine SPARAM(PS,P,IC)
      Implicit None

      Include 'globals.inc'

      Real*8 PS(MaxNSECP), P(MaxNSP+MaxNDE), IC(MaxNDE)

CC
C------------------------------------------------------------------C
C 7. Enter Equations Defining Secondary Paramters                  C
C        { e.g. PS(1) = P(1)*P(2) }                                C
C----c-------------------------------------------------------------C

      PS(1) = P(1)/P(2)
      PS(2) = PS(1)
      If(PS(2).ne.0.0) PS(3) = DLOG(2.0D0)/PS(2)

C------------------------------------------------------------------C
C------------------------------------------------------------------C
C
      Return
      End
```

APPENDIX 12.2   S-PLUS SCRIPT TO CREATE NONMEM DATASET TEMPLATES

```
#####   Create   data   template   for   4-samples/subj   designs
   ###################################
rm(.Random.seed)
set.seed <- 555
nSubj <- 200


############## Assign Nominal Values to reserved NONMEM Variables
   #######################


##### Specify time vector of dosing and obs events
# Specify dosing times for a subject
Dosetime.v <- c(0, 0, 12, 24, 36, 48)
# Specify obs times for a subject
Obstime.v <- c(0.5, 4, 48+2.5, 48+12)
# Combine dosing and obs times
time.v <- c(Dosetime.v, Obstime.v)
# Specify sampling time point # (value=0 for dosing records)
snum.v <- c(rep(0, length(Dosetime.v)), seq(1,length(Obstime.v)))


##### Specify EVID vector corresponding to time.v
EVID.v <- c(rep(1, length(Dosetime.v)), rep(0, length(Obstime.v)))


##### Specify AMT vector for a subject . . .
# . . . a single 30 mg (IV) loading dose, followed by 84.6 mg (PO) BID
amt.v <- c(30, rep(84.6, length(Dosetime.v)-1))
# Pad with missing values for obs records
amt.v <- c(amt.v , rep(NA, length(Obstime.v)))


##### Specify CMT vector for a subject
# Dose compartments
cmt.v <- c(2, rep(1, length(Dosetime.v)-1))
# Pad with obs compartments
cmt.v <- c(cmt.v, rep(2, length(Obstime.v)))


##### Specify dataset particulars
nrowSubj <- length(time.v) #(6 dose recs + 4 obs recs)

# Specify names of NONMEM input vars
simDataNames <- c("SUBJ", "TIME", "EVID", "AMT", "DV", "CMT", "SNUM")

simData <- data.frame(matrix(data=NA, nrow=nSubj*nrowSubj,
   ncol=length(simDataNames),
                     dimnames=list(NULL,simDataNames)))

# Assign Subj Nos.
simData$SUBJ <- rep(1000+seq(1:nSubj), each=nrowSubj)
```

```
##### Assign vectors to simData
simData$TIME <- rep(time.v, nSubj)
simData$EVID <- rep(EVID.v, nSubj)
simData$AMT <- rep(amt.v, nSubj)
simData$CMT <- rep(cmt.v, nSubj)
simData$SNUM <- rep(snum.v, nSubj)


# Reorder simData


simData <- simData[order(simData$SUBJ, simData$TIME, -simData$EVID), ]


##### Replace Nominal (PFIM-OPT) sample timepoints with randomly
   assigned timepoints
##### Method: Informative_Randomized_Block_1
simData.IBR1 <- simData
##### Replace nominal time by randomly selecting from a vector of
   pre-specified timepoints
# Specify the vector of times for each nominal timepoint
IBR1.lst <- list("samp1"=c(0.25, 0.5, 0.75, 1),
                 "samp2"=c(2, 3, 4, 6, 8),
                 "samp3"=48 + c(2, 4, 6),
                 "samp4"=48 + c(8, 10, 12))
#
for (iSnum in 1:length(Obstime.v)){
   TF.v <- simData.IBR1$SNUM==iSnum
   simData.IBR1$TIME[TF.v] <- sample(IBR1.lst[[iSnum]], size=sum(TF.
   v), replace=T)
}


##### Method: Informative_Randomized_Block_2
simData.IBR2 <- simData
##### Replace nominal time by randomly selecting from a uniform
   distribution
# Specify the min and max for uniform distributions for each nominal
   timepoint
IBR2.lst <- list("samp1"=c(0.25, 1.0),
                 "samp2"=c(3, 5),
                 "samp3"=48 + c(1.5, 3.5),
                 "samp4"=48 + c(10, 12))
#
for (iSnum in 1:length(Obstime.v)){
   TF.v <- simData.IBR2$SNUM==iSnum
   simData.IBR2$TIME[TF.v] <-
                  round(runif(sum(TF.v), IBR2.lst[[iSnum]][1],
   IBR2.lst[[iSnum]][2]),2)
}

############### Export Simulation Data Templates ###############
```

```
NMdataDir <- "./NMdatasets/"
options(digits=5)

TF.export <- F
if(TF.export){
    fileName <- paste(NMdataDir, "SimTmpl_PFIM4.csv", sep=")
    z.exportNM(simData, fileName)

    fileName <- paste(NMdataDir, "SimTmpl_IBR4A.csv", sep=")
    z.exportNM(simData.IBR1, fileName)
#
    fileName <- paste(NMdataDir, "SimTmpl_IBR4B.csv", sep=")
    z.exportNM(simData.IBR2, fileName)
}
```

APPENDIX 12.3   NONMEM CONTROL FILE

Simulate Clinical Trials for a Design Specified by the NONMEM Datafile Template Given by SimTmpl_Design.csv

```
$PROB ENOXAPARIN PK

$INPUT ISIM ID=SUBJ TIME EVID AMT DV CMT

$DATA ../ SimTmpl_Design.csv IGNORE=#
; Replace SimTmpl_Design.csv by appropriate simulation dataset template

$SUBROUTINE ADVAN2, TRANS2

$PK
; Specify Fixed Effect Parameters
   ACL = THETA(1)
   AV2 = THETA(2)
   AKA = THETA(3)

; Specify IIV Random Effect Parameters
   ZACL = ETA(1)
   ZAV2 = ETA(2)
   ZAKA = ETA(3)

; ####################################################
; Specify Individual PK Parameter Models

   CL = ACL*EXP(ZACL)
   V = AV2*EXP(ZAV2)
   KA = AKA*EXP(ZAKA)

   S2=V/100 ; 1 mg Enox is equiv to 100 IU of anti-Factor Xa activity
```

```
; Record simulated trial replicate
   ISIM = IREP

$ERROR
   IPRE = F
   IRES = DV - IPRE
   WFAC = F
   IWRE = IRES/WFAC
   Y = IPRE + WFAC*EPS(1)

$THETA
   0.708 ; ACL [L/hr]
   5.49 ; AV2 [L]
   0.232 ; AKA [1/hr]

$OMEGA
0.175 ; ZCL
0.277 ; ZV2
0 FIXED ; ZKA

$SIGMA
0.0682 ; PERR

$SIMULATION (ZZZZZ)

$TABLE ISIM SUBJ TIME EVID AMT DV CMT
        NOPRINT NOHEADER FILE=SimDesign.tab
```

APPENDIX 12.4   UNIX SHELL SCRIPT TO CREATE MULTIPLE SIMULATION FILES

```
#!/bin/csh
# Script Name: gen_rctl.csh

##### Generate multiple simulation files
# Syntax: gen_rctl nIter ctl_file
# where,
# nIter . . . number of simulation files to be generated
# ctl_file . . . name of simulation control file

set nIter=$1
set ctl_file=$2

set iter=1

while ($iter <= $nIter)
   set iter_ctl_file=${ctl_file:r}.{$iter}.ctl
   echo -n "$iter_ctl_file, Random: "
```

```
    set rnum=`random.bash`
    echo $rnum

# Replace random seed placeholder in simulation file "ZZZZZ"
# by a random seed generated by the operating system
sed s/ZZZZZ/$rnum/ $ctl_file > tmp.txt

# Replace "SimDesign" in the line below by . . .
# . . ."SimPFIM4" or "SimIBR4A", or "SimIBR4B" as applicable
sed s/SimDesign/SimDesign.$iter/ tmp.txt > $iter_ctl_file

    @ iter++
end

rm tmp.txt
```

APPENDIX 12.5   NONMEM CONTROL FILE

Estimate Parameters for Simulated Data Given in SimDesign.tab

```
$PROB ENOXAPARIN PK

$INPUT ISIM ID=SUBJ TIME EVID AMT DV CMT

$DATA ../SimDesign.csv IGNORE=#

; Replace SimDesign.csv by SimPIFM.csv, SimIBR4A.csv, or SimIBR4B.csv

$SUBROUTINE ADVAN2, TRANS2

$PK
; Specify Fixed Effect Parameters
   ACL = THETA(1)
   AV2 = THETA(2)
   AKA = THETA(3)

; Specify IIV Random Effect Parameters
   ZACL = ETA(1)
   ZAV2 = ETA(2)
   ZAKA = ETA(3)

; #######################################################
; Specify Individual PK Parameter Models

   CL = ACL*EXP(ZACL)
   V = AV2*EXP(ZAV2)
   KA = AKA*EXP(ZAKA)
```

```
   S2=V/100 ; 1 mg Enox is equiv to 100 IU of anti-Factor Xa activity

; Record simulated trial replicate
   ISIM = IREP

$ERROR
   IPRE = F
   IRES = DV - IPRE
   WFAC = F
   IWRE = IRES/WFAC
   Y = IPRE + WFAC*EPS(1)

$THETA
   1 ; ACL [L/hr]
   5 ; AV2 [L]
   0.2 ; AKA [1/hr]

$OMEGA
   0.2 ; ZCL
   0.3 ; ZV2
   0 FIXED ; ZKA

$SIGMA
   0.1 ; PERR

$EST SIG=5 PRINT=0 MAXEVAL=9000 NOABORT METHOD=1 INTER
```

APPENDIX 12.6   UNIX SHELL SCRIPT TO CREATE MULTIPLE ESTIMATION FILES

```
#!/bin/csh

# Script Name: gen_ectl.csh

##### Generate NM control files from a template file by changing the
#     input data filename from SimDesign to SimDesign.####,
#     where #### is iteration number
#
# Syntax: gen_rctl nIter ctl_file
# where,
# nIter . . . number of simulation files to be generated
# ctl_file . . . name of simulation control file

set iter=1

while ($iter <= $nIter)
# Change "SimDesign" in line below to "SimPFIM" or "SimIBR4A", or
   "SimIBR4B"
```

```
 seds/SimDesign/SimDesign.$iter/$ctl_file>${ctl_file:r}.{$iter}.ctl
 echo ${ctl_file:r}.{$iter}.ctl
 @ iter++
end
```

APPENDIX 12.7   PERL CODE TO EXTRACT PARAMETERS ESTIMATES FROM NONMEM OUTPUT FILES

```perl
#!/usr/bin/perl -w
#
# Program to extract NONMEM parameter estimates from multiple runs . . .
# . . . one set of estimates per line
# Syntax: extract_results.1line.pl *.lst
#         where "*.lst" is wildcard for selecting multiple NONMEM
#              result files
# Output: The extracted parameter estimates are output to . . .
#         the terminal (standard output)
#          and can be redirected to a summary result file using
#              "> summary.res"
#

foreach $pathname (@ARGV){
   $start = rindex($pathname,'/');
   $end = length($pathname);
   if($start < 0){ # then there is no "/" in the $pathname
   $filename = $pathname;
   }
   else{
   $filename = substr($pathname, $start+1, $end-$start-1);
   }
   $end = index($filename,".lst");
# $runID = substr($filename,0,$end);

   open(RESULT_FILE,$pathname);
   @input_lines = <RESULT_FILE>;

#@input_lines = <STDIN>;
$n_lines = @input_lines;

for ($jj=0; $jj <= $n_lines-1; $jj++){

   if($input_lines[$jj] =~/PROBLEM NO/){
#     printf("%s ",$runID);
      }

   if($input_lines[$jj] =~/MINIMIZATION/){
      @estMsgLine = split /\s+/, $input_lines[$jj];
      $estMsg = $estMsgLine[1];
```

```
        printf("%s ",$estMsg);
        }

##### Extract the Objective Function Value
     if($input_lines[$jj] =~ /MINIMUM VALUE OF OBJECTIVE FUNCTION/){
   @objFunLine = split /\s+/, $input_lines[$jj+9];
        $objFunValue = $objFunLine[2];
   printf("%9.4f ",$objFunValue);
#   printf("\n FINAL PARAMETER ESTIMATES ");

##### Extract the THETA variables
   @theta = split /\s+/, $input_lines[$jj+24];
        $len_theta = @theta;
#        printf("\n %s ", "THETAS:");
         for($ii=1; $ii <= $len_theta-1; $ii++){
        printf("%s ", $theta[$ii]);
   }

##### Extract the Eta variables (currently does not handle COVARIANCE)
# Find out how many Eta variables are in output
        $etaLine = $input_lines[$jj+31];
#         printf("\n %s ", "ETAS: ");
        $countEta = 0;
   while($etaLine =~ /ETA/g){
        $countEta++;
   }

# Initialize the @eta array
@eta = (0) x $countEta;

# Extract Eta variables from multiple lines
   for($iEta=1; $iEta <= $countEta; $iEta++){
         @etaLine = split /\s+/, $input_lines[$jj+31+3*$iEta];
      $eta[$iEta] = $etaLine[$iEta];
      printf("%s ", $eta[$iEta]);
}

##### Extract the Sigma variables
# Locate the line at which the SIGMA variable starts
        $ii = 31;
        while($input_lines[$jj+$ii] !~ /SIGMA/){
        $ii++;
   }

# Find out how many Sigma variables are in output
        $sigmaLine = $input_lines[$jj+$ii+3];
#         printf("\n %s ", "SIGMAS:");
        $countEps = 0;
   while($sigmaLine =~ /EPS/g){
```

```
        $countEps++;
    }


# Initialize the @eps array
@eps = (0) x $countEps;


# Extract Eps variables from multiple lines
    for($iEps=1; $iEps <= $countEps; $iEps++){
            @epsLine = split /\s+/, $input_lines[$jj+$ii+3+3*$iEps];
            $eps[$iEps] = $epsLine[$iEps];
            printf("%s ", $eps[$iEps]);
              }
              printf(" \n")
          }


    }
# printf(" \n")
    }
```

APPENDIX 12.8   S-PLUS CODE TO ANALYZE AND COMPARE NONMEM ESTIMATION RESULTS

```
##################### Read in orig Data
dataDir <- "./NMresults/"
plotDir <- "./Plots/"

export.plot <- T

fileName <- "Est.PFIM_4.100.res"
Est.PFIM.4sample <- read.table(paste(dataDir, fileName, sep=""),
header=F)
#
fileName <- "Est.IBR1_4.100.res"
Est.IBR1.4sample <- read.table(paste(dataDir, fileName, sep=""),
header=F)
#
fileName <- "Est.IBR2_4.100.res"
Est.IBR2.4sample <- read.table(paste(dataDir, fileName, sep=""),
header=F)


colNames <- c("M.STATUS", "OFV", "CL.TV", "V.TV", "KA.TV",
                        "CL.OM", "V.OM", "KA.OM", "ERR.VAR")

names(Est.PFIM.4sample) <- colNames
names(Est.IBR1.4sample) <- colNames
names(Est.IBR2.4sample) <- colNames
#
```

```
Est.PFIM.4sample$DESIGN <- rep("PFIM-4", nrow(Est.PFIM.4sample))
Est.IBR1.4sample$DESIGN <- rep("IBR-4A", nrow(Est.IBR1.4sample))
Est.IBR2.4sample$DESIGN <- rep("IBR-4B", nrow(Est.IBR2.4sample))
#

Est.all    <-    rbind(Est.PFIM.4sample,    Est.IBR1.4sample,    Est.
IBR2.4sample)

Est.all$DESIGN <- ordered(Est.all$DESIGN,
                          levels=c("PFIM-4", "IBR-4A", "IBR-4B"))

CL.TV.true <- 0.708
V.TV.true <- 5.49
KA.TV.true <- 0.232
#
CL.OM.true <- 0.175
V.OM.true <- 0.277
#
ERR.VAR.true <- 0.0682

#### Get Rel Error for Est.all
Est.all$CL.TV.RelErr <- (Est.all$CL.TV - CL.TV.true)/CL.TV.true * 100
Est.all$V.TV.RelErr <- (Est.all$V.TV - V.TV.true)/V.TV.true * 100
Est.all$KA.TV.RelErr <- (Est.all$KA.TV - KA.TV.true)/KA.TV.true * 100
#
Est.all$CL.OM.RelErr <- (Est.all$CL.OM - CL.OM.true)/CL.OM.true * 100
Est.all$V.OM.RelErr <- (Est.all$V.OM - V.OM.true)/V.OM.true * 100
#
Est.all$ERR.VAR.RelErr <- (Est.all$ERR.VAR - ERR.VAR.true)/ERR.VAR.
true * 100

Est.all$STATUS.FLAG <- Est.all$M.STATUS=="SUCCESSFUL"
Success.v <- tapply(Est.all$STATUS.FLAG, Est.all$DESIGN, sum)

Median.CL.TV.RelErr <- tapply(Est.all$CL.TV.RelErr, Est.all$DESIGN,
  median)
Median.V.TV.RelErr <- tapply(Est.all$V.TV.RelErr, Est.all$DESIGN,
  median)
Median.KA.TV.RelErr <- tapply(Est.all$KA.TV.RelErr, Est.all$DESIGN,
  median)
Median.CL.OM.RelErr <- tapply(Est.all$CL.OM.RelErr, Est.all$DESIGN,
  median)
Median.V.OM.RelErr <- tapply(Est.all$V.OM.RelErr, Est.all$DESIGN,
  median)
Median.ERR.VAR.RelErr  <-  tapply(Est.all$ERR.VAR.RelErr,  Est.
  all$DESIGN, median)

Median.CL.TV.RelErr  <-  tapply(abs(Est.all$CL.TV.RelErr),  Est.
  all$DESIGN, median)
```

```
Median.V.TV.RelErr    <-    tapply(abs(Est.all$V.TV.RelErr),    Est.
   all$DESIGN, median)
Median.KA.TV.RelErr   <-    tapply(abs(Est.all$KA.TV.RelErr),    Est.
   all$DESIGN, median)
Median.CL.OM.RelErr   <-    tapply(abs(Est.all$CL.OM.RelErr),    Est.
   all$DESIGN, median)
Median.V.OM.RelErr    <-    tapply(abs(Est.all$V.OM.RelErr),    Est.
   all$DESIGN, median)
Median.ERR.VAR.RelErr <- tapply(abs(Est.all$ERR.VAR.RelErr), Est.
   all$DESIGN, median)

############################  Plots    of    Analysis    Results
    ################################

###### Box Plots of %Bias
graphsheet(orientation="p")
par(mfrow=c(3,1))
par(mar=c(4,4,1,2)+.1)
par(mgp=c(2,0.75,0))
#
boxplot(split(Est.all$CL.TV.RelErr, Est.all$DESIGN),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-25, 25))
abline(h=0)
mtext("CL.TV", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(Est.all$V.TV.RelErr, Est.all$DESIGN),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-50, 50))
abline(h=0)
mtext("V.TV", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(Est.all$KA.TV.RelErr, Est.all$DESIGN),
     names=paste(as.character(unique(Est.all$DESIGN)),"\
 n(",Success.v, "%)", sep=""),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-50, 50))
abline(h=0)
mtext("KA", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(Est.all$CL.OM.RelErr, Est.all$DESIGN),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-50, 50))
abline(h=0)
mtext("CL.OM", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)
#
#
boxplot(split(Est.all$V.OM.RelErr, Est.all$DESIGN),
```

```
      confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-50, 50))
abline(h=0)
mtext("V.OM", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(Est.all$ERR.VAR.RelErr, Est.all$DESIGN),
      names=paste(as.character(unique(Est.all$DESIGN)),"\
  n(",Success.v, "%)", sep=""),
      confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(-50, 50))
abline(h=0)
mtext("ERR.VAR", line=0.5)
mtext("Bias [%]", side=2, outer=F, line=2, cex=1)

#mtext("Summary of Estimation Results (METHOD=FOCEI)", side=3,
outer=T, line=-2, cex=1)

if(export.plot){
  fileName <- paste(plotDir, "Bias###",".wmf",sep="")
  export.graph(fileName,Name=paste("GSD",dev.cur(),sep=""),ExportType
  ="WMF")
}

##### Box Plots of %Precision
graphsheet(orientation="p")
par(mfrow=c(3,1))
par(mar=c(4,4,1,2)+.1)
par(mgp=c(2,0.75,0))
#
boxplot(split(abs(Est.all$CL.TV.RelErr), Est.all$DESIGN),
      confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 25))
abline(h=25, lty=2)
mtext("CL.TV", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)
#
#
boxplot(split(abs(Est.all$V.TV.RelErr), Est.all$DESIGN),
      confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 50))
abline(h=25, lty=2)
mtext("V.TV", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(abs(Est.all$KA.TV.RelErr), Est.all$DESIGN),
      names=paste(as.character(unique(Est.all$DESIGN)),"\
  n(",Success.v, "%)", sep=""),
      confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 50))
abline(h=25, lty=2)
mtext("KA", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)
```

```
#
boxplot(split(abs(Est.all$CL.OM.RelErr), Est.all$DESIGN),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 50))
abline(h=25, lty=2)
mtext("CL.OM", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(abs(Est.all$V.OM.RelErr), Est.all$DESIGN),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 50))
abline(h=25, lty=2)
mtext("V.OM", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)
#
boxplot(split(abs(Est.all$ERR.VAR.RelErr), Est.all$DESIGN),
     names=paste(as.character(unique(Est.all$DESIGN)),"\
 n(",Success.v, "%)", sep=""),
     confint=T, confnotch=T, confcol=0, medcol=1, ylim=c(0, 50))
mtext("ERR.VAR", line=0.5)
mtext("Absolute Error [%]", side=2, outer=F, line=2, cex=1)

if(export.plot){
  fileName <- paste(plotDir, "Precision###",".wmf",sep="")
  export.graph(fileName,Name=paste("GSD",dev.cur(),sep=""),ExportType
  ="WMF")
}

#mtext("Summary of Estimation Results (METHOD=FOCEI)", side=3,
  outer=T, line=-2, cex=1)
```

Population Models for Drug Absorption and Enterohepatic Recycling

OLIVIER PÉTRICOUL, VALÉRIE COSSON, ELIANE FUSEAU, and
MATHILDE MARCHAND

13.1 INTRODUCTION

With the emergence of biotechnologies and the biomolecular techniques more than 20 years ago, the screening of new chemical entities (NCEs) has been focused more on receptor binding properties than administration–distribution–metabolism–excretion (ADME) discrimination. This has led to the development of compounds with poor ADME properties. Among the pharmacokinetic (PK) properties of a NCE, a low and highly variable bioavailability results in poorly controlled plasma concentrations and drug effects. This is often the main reason for the failure of the development of a compound (1–4) and high attrition rates. In that sense, Lipinsky (5) recommends that the order of testing (pharmacological activity versus ADME properties) may change in order to discriminate early the NCE with poor PK properties. Nevertheless, significant numbers of drugs currently under development have physicochemical and/or PK properties that are less than ideal for the oral route, which is often favored.

With the development of new formulation technologies, bioavailability may be improved when a compound has poor bioavailability. Metabolism inhibitors, prodrugs, membrane permeation enhancers, ion pairing and complexation, and particle carriers are examples of strategies to improve bioavailability (6). Formulation could also change during development or during the lifetime of a drug, in order to allow for new dosing regimens, line product extension, or strong patients' needs. Matrix tablet, osmotic tablet, and particle coating are examples of formulation changes that can occur.

Therefore, two different situations can be highlighted for orally administered compounds. On one hand, the compound itself is hard to deliver, due to poor physicochemical and/or ADME properties; and on the other hand, the absorption of the compound is "controlled" by formulation properties. In both situations, the absorption profile is called atypical (or irregular absorption profile), in comparison

with a typical absorption profile, where the absorption of the compound follows a first-order or zero-order kinetic (7, 8).

In this chapter, we examine the analysis of both typical and atypical absorption profiles, along with enterohepatic recycling (EHR), which could affect both the bioavailability and the absorption profile of a drug in a population of subjects.

13.2   GENERAL CONSIDERATIONS

13.2.1   Scope and Definitions

The oral route is most preferred for drug therapy. Other extravascular routes, such as subcutaneous, intramuscular, intranasal, lung delivery, and transdermal, also present some advantages in drug therapy. This chapter focuses on oral absorption, but models presented apply to other extravascular routes as well.

The terms absorption and bioavailability are often used interchangeably, meaning that the same definition may apply to both, that is, the degree to which or the rate at which a drug is absorbed or becomes available at the site of measurement (or of action) after extravascular administration. For the purpose of this chapter, absorption is defined as the transfer from the site of administration across biologic barriers to a site where it is measured (e.g., the blood). Bioavailability is defined as the fraction of a dose administered that reaches the general circulation or the site of action (9). It is relatively important to distinguish these two parameters in an absorption model, and the omission of the rate of absorption may have serious consequences on the interpretation of bioavailability depending on the underlying model. For example, in the case of endogenous substances, the term "bioavailability" is ambiguous unless one specifies whether it refers to availability of the exogenous substance only or the sum total of the exogenous and endogenous substances.

Although intravenous (IV) dosing is very important for the determination of both clearance and volume of distribution, it is not addressed in this chapter. Therefore, the analysis of the absolute bioavailability of a substance is not addressed. However, relative bioavailability (food effect, drug–drug interaction, time effect, dose nonproportionality, etc.) is covered.

Therefore, two different sets of parameters are defined:

- Parameters describing the rate of absorption
- Parameters describing the amount absorbed

Noncompartmental analysis (NCA) is the most frequently used method and provides good information about the absorption rate. For example, the concept of partial area under the curve (AUC) has been evaluated in comparative PK studies, and these metrics had greater statistical power than the peak plasma drug concentrations (C_{max}) (10). However, NCA requires more samples than are customarily available in Phase 2/3 studies.

In this chapter, we focus on both the rate of absorption (e.g., rate constant k_a for a first-order kinetic profile) and the bioavailability (usually represented by the letter F).

13.2.2   Context: When Does One Need to Describe Absorption?

There is often a need to evaluate clinically significant PK interactions in terms of rate and extent of absorption, by estimating C_{max}, time to C_{max} (t_{max}), and AUC. These are considered important variables for describing exposure. Rate of absorption can be increased, more often decreased, resulting in changes of t_{max} and C_{max}; or the extent of absorption can be increased or decreased, affecting both C_{max} and AUC (11). When efficacy and/or safety can be related to C_{max} in plasma or in another tissue or compartment, it may be important to describe accurately the absorption phase after single or repeat dose administration (12).

In acute disease conditions, such as pain, migraine, or emesis (13, 14), the rapidity of onset of action is of primary importance. Delayed absorption of sumatriptan, due to encapsulation of the tablet in comparative studies, may account for the lower efficacy of sumatriptan in comparative studies (15). On the contrary, a delayed absorption of furosemide, when it is prescribed in patients with congestive heart failure, on a chronic use may result in diuretic resistance (16, 17). The occurrence of adverse events may be related to C_{max}, as for fluoroquinolones that present cardiotoxicity (18); while adverse events for moxifloxacin used in clinical trials as a positive control for Q-T interval studies (19, 20) may be related to concentration in plasma rather than to AUC. High plasma concentrations of acyclovir may lead to precipitation of the drug in renal tubules due to low solubility, resulting in renal tubular damage and acute renal failure (21, 22). The pharmacokinetic/pharmacodynamic (PKPD) modeling of diltiazem, a calcium channel blocker, shows absorption rate dependency of the hysteresis loop (23, 24). The maximum effects of sildenafil on blood pressure and heart rate occur at peak plasma concentrations (25, 26).

On the contrary, the absorption of some drugs used in chronic diseases does not appear relevant to predict their efficacy, as for antiepileptic drugs (27) and antiretroviral drugs such as antiproteases (28, 29). For digoxin, most of the models linked the time course of positive inotropic effect to digoxin amount in peripheral or effect compartment (30, 31), and, therefore, absorption phase does not present a real interest and is in fact irrelevant.

Table 13.1 presents some examples of drugs where absorption characteristics may or may not be important and suggests whether modeling the absorption is necessary for the development of the compound or not.

13.2.3   Sources of Variability in Absorption

The oral absorption of drugs is an extremely complex phenomenon that manifests itself through the interaction between drug and patient-specific variables, as depicted in Figure 13.1.

Among drug properties that impact absorption, three are noteworthy:

- Physicochemical properties of the compound such as pK_a, solubility, and lipophilicity
- Formulation characteristics such as the particle size, surface area, crystal form, and dosage forms (solution, tablet, capsule, suspension, emulsion, gel, and modified released)
- Pharmacological drug properties regrouping the drug affinity for transporters and enzymes of metabolism

TABLE 13.1   When Should We Model the Absorption? A Review of Some Examples

Characteristic	Absorption Modeling	Acute Treatment	Chronic Treatment
Efficacy	+	Analgesic: morphine (14) Antimigraine: triptans (15) Antiemetic: ondansetron (13) Erectile dysfunction: sildenafil (26)	Diuretic: furosemide (16, 17)
	−		Antiepileptic (27) Antiretroviral drug (28, 29)
Safety	+	Erectile dysfunction: sildenafil (25, 26) Antimigraine: triptans (15)	Antiviral drug: valaciclovir (21, 22) Antibiotic: moxifloxacin (19–20) Calcium channel blocker: diltiazem (23, 24)
	−		Digitalis drug: digoxin (30, 31)
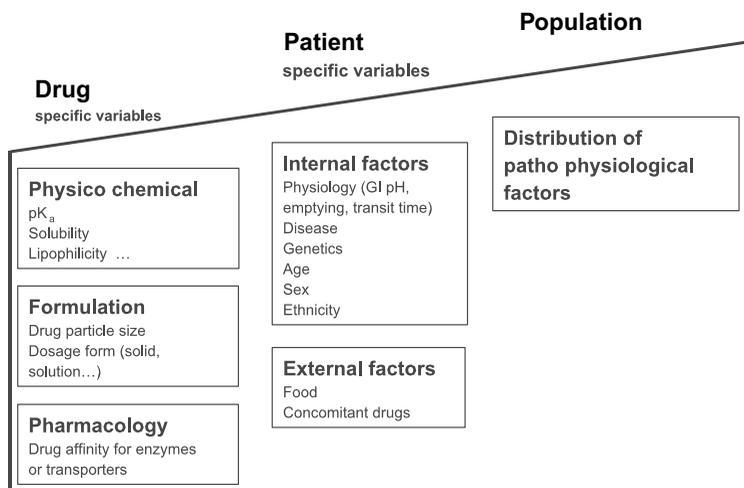
FIGURE 13.1   Variables to take into account in population absorption model.

The impact of the physicochemical properties of a compound on rate and extent of drug absorption has been extensively reported (5, 32–35). For example, the sensitivity of absorption to particle size decreases with increasing dose and solubility (36) and formulation characteristics affect the dissolution rate and, subsequently, rate and extent of absorption (37).

Absorption can be affected by the presence of an efflux mechanism such as P-glycoprotein. Drug binding may affect oral absorption: fluoroquinolones bind to cations and form insoluble chelates, resulting in a decrease of bioavailability (11).

Usually, investigation of these properties is performed during drug discovery or at preclinical stage for lead compound selection (37, 38). In this chapter, an effort is made to describe absorption as an overall process, and not to describe each mechanism.

Among the patient-specific variables affecting absorption, one can differentiate between internal and external factors. Internal factors include gastrointestinal (GI) tract function, represented by pH, stomach emptying time, and transit time varying with age, gender, and diseases (Crohn's disease, celiac disease, AIDS enteropathy, drug- and irradiation-induced malabsorption) (7). Other sources of within-subject variability in absorption include diurnal factors, changes in blood flow, body position, and volume of fluid intake. External factors include food, alcohol, or concomitant medications that may affect the dissolution of the drug or GI function.

In early drug development, the description of the absorption profile is important for the selection of the most suitable lead compound. In later clinical phases of full development, the description of the variability in drug absorption may become more important, for evaluation of safety and efficacy.

The population pharmacokinetic (PPK) modeling approach should be executed while taking into account the physicochemical properties of a drug, the pathophysiology of a patient, and the variability of all the different mechanisms of absorption.

The description and quantification of between-subject variability become very important in a population of patients. For example, the double-peak phenomenon observed on median PK profiles of sustained-release diclofenac is due to large differences in individual t_{max} values (39).

13.2.4 The Data: Experimental Design for Assessing Drug Absorption and Enterohepatic Recycling

Intensive and detailed Phase 1 and 2 studies of small groups of healthy subjects or patients provide the most complete picture of a drug's essential properties. The following variables must be collected and accounted for in population modeling of the absorption process.

- *A full drug dose and concentration data* profile with or without IV data to identify a structural model with adequate absorption phase samples to estimate the parameters must be collected. Absorption is best described by physiologically based modeling. However, development of such a model necessitates a large amount of data, which is seldom available in humans. Use of animal studies, radiolabels (usefulness is limited as not only parent drug but also metabolites are measured if there is presystemic metabolism), Caco-2, Madin–Darby canine kidney (MDCK) cell based, parallel artificial membrane permeation assay (PAMPA), isolated segment of gastrointestinal tract (animal), perfusion methods (segment of the gut is perfused and cannulated), and Ussing chamber

may provide such data. If the influence of the absorption rate needs to be evaluated in patients, then one needs to design a sampling protocol to allow such estimation.

- *All the factors affecting absorption* as presented in Figure 13.1 should be addressed. Salient data need to be collected and patients monitored. Patients may need to be stratified for various factors such as concomitant medications, sex, and age.
- *Blood sampling strategies* may need to be selected in relation to meal times to describe enterohepatic recycling.
- *One must account for PK/PD data* so that if one wants to discriminate between the influence of C_{max} or AUC on clinical efficacy/safety markers, special attention should be paid to experimental design (i.e., optimizing for both PK samples and PD assessments).

13.2.5 Analysis Strategy for Evaluating a Suitable Population Absorption Model

Rich Data: Structural Model

- *Plots.* A general procedure has been proposed for the analysis of absorption profiles following oral administration of a drug (7, 8), where it is strongly suggested to plot first concentration–time data on both linear and log-linear scales by subject before attempting to model the data. Visual examination of the graphs helps to identify whether the absorption profiles are typical or atypical. This allows one to choose the most suitable absorption model for the majority of the patients. For large populations, "spaghetti" plots are preferred.
- *Individual Analysis.* However, if the plots are too different between subjects and no obvious absorption model is suggested by the graphs, one should start with individual modeling on a range of doses and demographic groups representative of the population.
- *Secondary Peaks.* If the plots display secondary peak, the AUC represented by the second peak should be estimated in relation to total AUC to decide if the model needs to describe this peak. With repeated secondary peaks, intravenous (IV) data are critical to assert the presence of enterohepatic recycling.
- *Deconvolution.* The absorption profile may be represented by plots of instantaneous rate of absorption as a function of time, through deconvolution (8).

Sparse Data, Many Subjects: Population Model

- *Sampling Design Issues.* Often, in Phase 3 clinical studies, the absorption kinetics are difficult to characterize with precision, because too few samples are taken during the early phase of absorption. In the case where plasma concentrations have already reached their highest values in the first samples, two approaches can be used. Either it is assumed that the drug was administered as a bolus IV or a constant rate infusion is defined up to the first plasma sample (40). It should be noted that fixing absorption parameters using prior knowledge could lead to biased estimates of disposition and elimination parameters.

- *Effect of Time and Repeated Dosing on Pharmacokinetics.* The majority of absorption models deal with single-dose administration. However, upon chronic administration of multiple doses of a drug, the time to reach the peak plasma concentration (40) or the maximum of plasma concentrations may change. Both rate and extent of absorption can vary over the duration of a study due to between-occasion variability or the time effect on the absorption of the drug.
- *Difficulties in Estimating Absorption Parameters.* Very often, little attention is given to sampling design in long-term Phase 2/3 studies. Identifiability of the population absorption and enterohepatic models need to be carefully assessed before launching into a full scale analysis of data (41). Identifiability problems may arise due to a high correlation between parameters (e.g., central volume of distribution and k_a) or collinearity (the effect of common covariates on several PK parameters). Moreover, a design based on healthy subjects data may prove inappropriate in patients who have a different absorption profile.

13.3   DRUG ABSORPTION MODELS

The typical absorption profiles are represented by the first-order absorption and the zero-order absorption. Atypical absorption profiles can be described by parallel first-order absorption, mixed first-order and zero order absorption, or Weibull-type absorption.

In most papers, a small number of different models are used. Those are described in detail here. All models are defined using differential equations, although, for the simple cases, analytical solutions exist for the description of plasma drug concentrations that obviously render the analyses quicker.

The models are illustrated using simulations of plasma concentrations of a drug X, in 50 subjects following single administration of a dose of 10 mg. For simplicity and clarity of the results, the drug X is assumed to follow a one-compartment disposition model, with population apparent clearance (*CL*) of 5 L/h, and population apparent volume of distribution (*V*) of 50 L. Between-subjects variability (BSV) is modeled using a proportional error model and is expressed as coefficient of variation (%CV). For both apparent clearance and volume of distribution, BSV is defined as 20% CV. Population parameters (fixed and random effects) for the different absorption models are presented in Table 13.2. Again, for simplicity and clarity of the absorption models presented, residual error is not taken into account in the simulations, although noise in the data is often a drawback to successfully estimate models of absorption data.

13.3.1   Example 1: First-Order Absorption Model

Oral drug absorption is often described as a first-order mechanism, and through compartmental modeling, oral absorption is represented by the first-order absorption rate constant, k_a (per time unit). Although it is not used in the current example, inclusion of lag time may be needed to better describe absorption processes. The kinetics of drug amount in the plasma following a first-order absorption process is described by a system of differential equations, as follows:

TABLE 13.2   Absorption Parameters Used for Simulations of One-Compartment Model with $CL = 5\,L/h$ (20% CV), $V = 50\,L$ (20% CV), and Dose $= 10\,mg$

Type of Absorption	Absorption Parameters	Values (Between-Subject Error)
First-order absorption	K_a (h^{-1})	0.5 (60% CV)
Zero-order absorption	Duration (h)	2 (60% CV)
Two parallel first-order absorptions	K_{a1} (h^{-1})	0.8 (60% CV)
	K_{a2} (h^{-1})	0.6 (60% CV)
	Lag time for second process (h)	5 (20% CV)
	Fraction of the dose absorbed through the first process	0.5 (20% CV)
Mixed first-order and zero-order absorption	K_a (h^{-1})	0.5 (60% CV)
	Duration (h)	2 (60% CV)
	Lag time for zero-order process (h)	2 (20% CV)
	Fraction of the dose absorbed through the first-order process	0.5 (20% CV)
Weibull absorption (one function)	K_a (h^{-1})	0.4 (60% CV)
	γ	4 (20% CV)
Weibull absorption (two functions)	K_{a1} (h^{-1})	2 (60% CV)
	$\gamma 1$	0.5 (20% CV)
	K_{a2} (h^{-1})	0.2 (60% CV)
	$\gamma 2$	4 (20% CV)
	Fraction of the dose absorbed through the first process	0.5 (20% CV)

$$\frac{dA(1)}{dt} = -k_a A(1)$$

$$(13.1)$$

$$\frac{dA(2)}{dt} = k_a A(1) - \frac{CL}{V} A(2)$$

$$(13.2)$$

with the initial conditions at time zero

$$A(1) = DOSE$$

$$A(2) = 0$$

and where k_a is the first-order absorption rate constant and CL and V are as defined above.

The plasma drug profiles in 50 subjects, following oral administration of the drug X and assuming a first-order absorption process, are plotted in Figure 13.2. Corresponding NONMEM control file and data set are in Appendix 13.1.

13.3.2   Example 2: Zero-Order Absorption Model

Although the assumption of first-order absorption is satisfactory for many drugs, the absorption of certain drugs is better described by assuming a zero-order (constant rate) absorption. The absorption can be described either by a constant rate
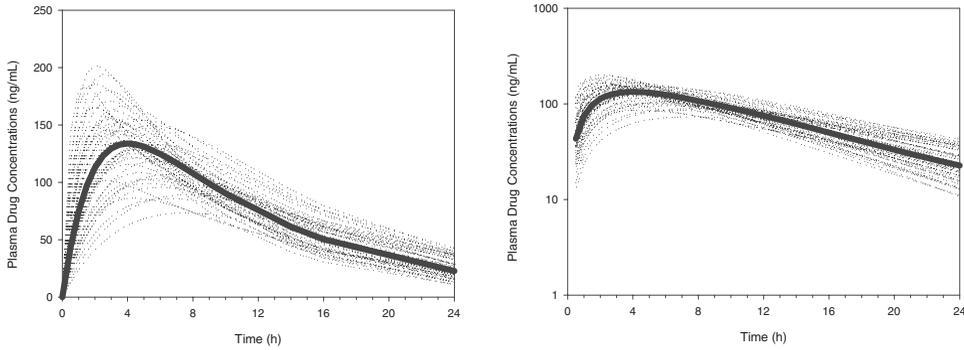
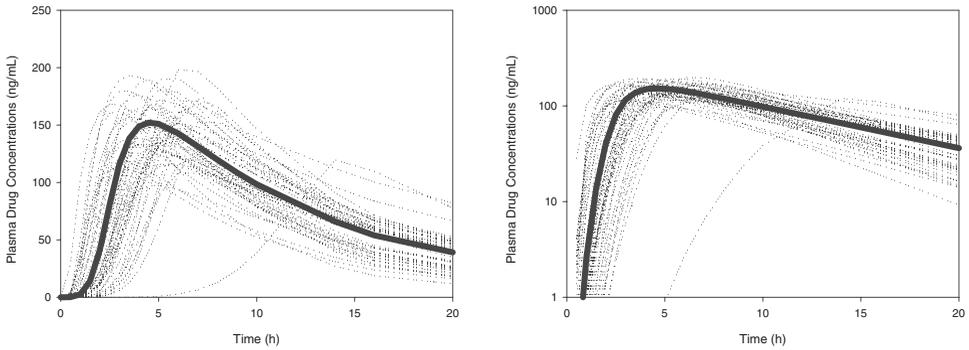**FIGURE 13.2** Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming a first-order absorption type. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.

(amount/time unit), or a duration of infusion (time unit). For the purpose of our example, the zero-order absorption is described by a duration $D1$, without a lag time. The kinetics of the drug amount in plasma is described by the following differential equations.

For time $\leq D1$,

$$\frac{dA(1)}{dt} = \frac{DOSE}{D1} - \frac{CL}{V}A(1) \qquad (13.3)$$

For time $> D1$,

$$\frac{dA(1)}{dt} = -\frac{CL}{V}A(1) \qquad (13.4)$$

with the initial condition at time zero

$$A(1) = 0$$

and where $D1$ is the duration of the zero-order absorption and CL and V are as defined above.

The plasma drug profiles in 50 subjects, following oral administration of the drug X and assuming a zero-order absorption process, are plotted in Figure 13.3. Corresponding NONMEM control file and data set are in Appendix 13.2.

13.3.3 Example 3: Two Parallel First-Order Absorption Models

In some cases, after oral administration, the plasma concentrations exhibit a double peak or shouldering-type absorption. One of the reasons to explain such kinetic behavior is the presence of an enterohepatic recycling. However for enterohepatic recycling, the double-peak appearance would be independent of the administration route and should be present after IV administration. Enterohepatic recycling is addressed in detail in Section 13.4.
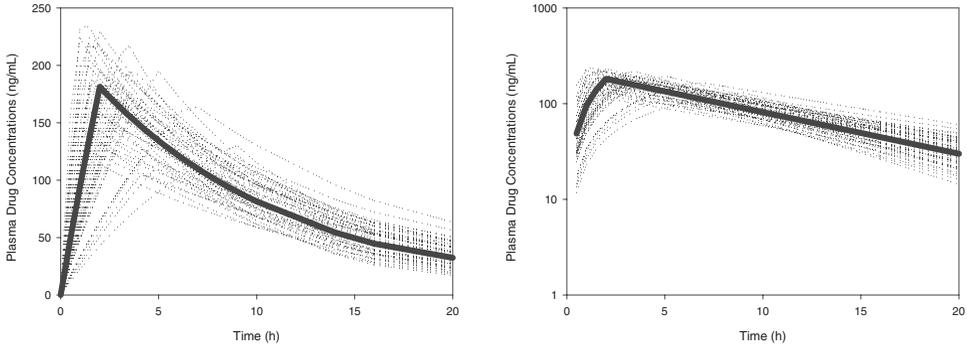
FIGURE 13.3 Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming a zero-order absorption type. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.

There are several explanations for the double-peak phenomenon after oral administration. There may be two distinct sites of absorption (42), gastric empty-ing limited absorption (43), or a variable gastric emptying rate (44–50), or it may be due to formulation characteristics (39). However, the concept of parallel first-order absorption is not limited to two absorption processes (51–55).

For example, a two parallel first-order absorption type can be one where the first process starts without a lag time and the second process starts with a lag time (ALAG2). Without lag time this absorption is called "simultaneous first-order absorption." In population modeling, the presence of more than one lag time often makes the model difficult to identify and in case the between-subject variability in lag time is important, there might be an identifiability problem for the population absorption model. Prior information, from preclinic or previous PK studies, may help in defining the structure of the model.

The kinetics of drug amount in the plasma following two parallel first-order absorp-tion processes is described by the following system of differential equations.

For time $\leq ALAG2$,

$$\frac{dA(1)}{dt} = -k_{a1}A(1) \tag{13.5}$$

$$\frac{dA(3)}{dt} = k_{a1}A(1) - \frac{CL}{V}A(3) \tag{13.6}$$

For time $> ALAG2$,

$$\frac{dA(1)}{dt} = -k_{a1}A(1) \tag{13.7}$$

$$\frac{dA(2)}{dt} = -k_{a2}A(2) \tag{13.8}$$

$$\frac{dA(3)}{dt} = k_{a1}A(1) + k_{a2}A(2) - \frac{CL}{V}A(3) \tag{13.9}$$
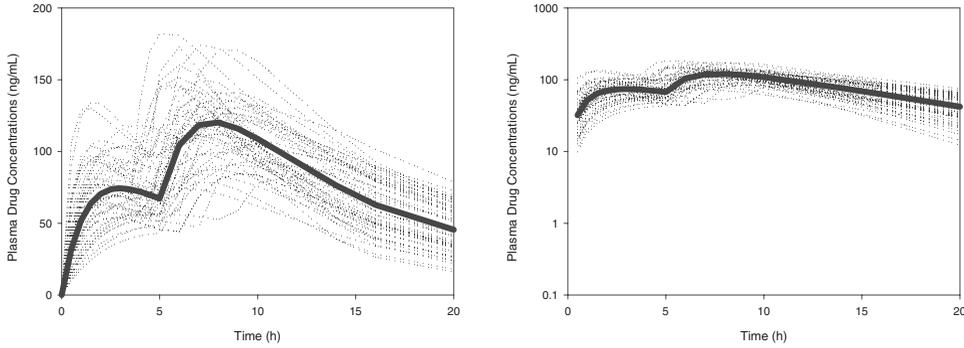
FIGURE 13.4 Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming two parallel first-order absorption types. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.

with the initial conditions at time zero

$$A(1) = f \cdot DOSE$$
$$A(2) = (1 - f) \cdot DOSE$$
$$A(3) = 0$$

and where k_{a1} is the rate constant of absorption for the first absorption process, k_{a2} is the rate constant of absorption for the second absorption process, f is the fraction of the dose absorbed through the first absorption process, $ALAG2$ is the lag time for the second absorption process to start, and CL and V are as defined above.

The plasma drug profiles in 50 subjects, following oral administration of the drug X and assuming two parallel first-order absorption processes, are plotted in Figure 13.4. Corresponding NONMEM control file and data set are in Appendix 13.3.

13.3.4 Example 4: Mixture of First-Order Absorption and Zero-Order Absorption Models

Sometimes, two first-order absorption processes do not adequately describe the data and the absorption profiles are better described by a combination of first-order and zero-order processes (40, 56–59). Lag time may be added for each type of absorption, which then will determine whether the two processes are simultaneous or sequential. Moreover, if the first-order rate constant is linked to the zero-order input parameters, the model can be interpreted as the consequence of dissolution-limited absorption. The ordering of the processes (first-order absorption first, or zero-order absorption first) is usually empirical or data driven. Pathophysiology and/or physicochemical characteristics of the compound may help in deciding the order.

For the current example, the first-order process starts immediately after dosing and is followed, with a lag time ($ALAG2$), by a zero-order process. The kinetics of drug amount in the plasma is described by a system of differential equations.

For time $\leq ALAG2$,

$$\frac{dA(1)}{dt} = -k_a A(1) \tag{13.10}$$

$$\frac{dA(2)}{dt} = k_a A(1) - \frac{CL}{V} A(2) \tag{13.11}$$

For $ALAG2 <$ time $\leq ALAG2 + D2$,

$$\frac{dA(1)}{dt} = -k_a A(1) \tag{13.12}$$

$$\frac{dA(2)}{dt} = k_a A(1) + \frac{(1-f) \cdot DOSE}{D2} - \frac{CL}{V} A(2) \tag{13.13}$$

For time $> D1 + ALAG2$,

$$\frac{dA(1)}{dt} = -k_a A(1) \tag{13.14}$$

$$\frac{dA(2)}{dt} = k_a A(1) - \frac{CL}{V} A(2) \tag{13.15}$$

with the initial conditions at time zero

$$A(1) = f \cdot DOSE$$

$$A(2) = 0$$

and where k_a is the rate constant of absorption for the first-order absorption process, $D2$ is the duration of absorption for the zero-order absorption process, f is the fraction of the dose absorbed through the first-order absorption process, $ALAG2$ is the lag time for the zero-order absorption process to start, and CL and V are as defined above.

The plasma drug profiles in 50 subjects, following oral administration of the drug X and assuming a mixture of first-order and zero-order absorption processes, are plotted in Figure 13.5. Corresponding NONMEM control file and data set are in Appendix 13.4.

13.3.5 Example 5: Weibull-Type Absorption Model

It may be the case that none of the above absorption models allows an adequate or appropriate description of the plasma concentration profiles. The use of Weibull function(s) may then provide an improved description of the data (60–66).

The kinetics of drug amount in the plasma following a Weibull absorption process is described by a system of differential equations, as follows:
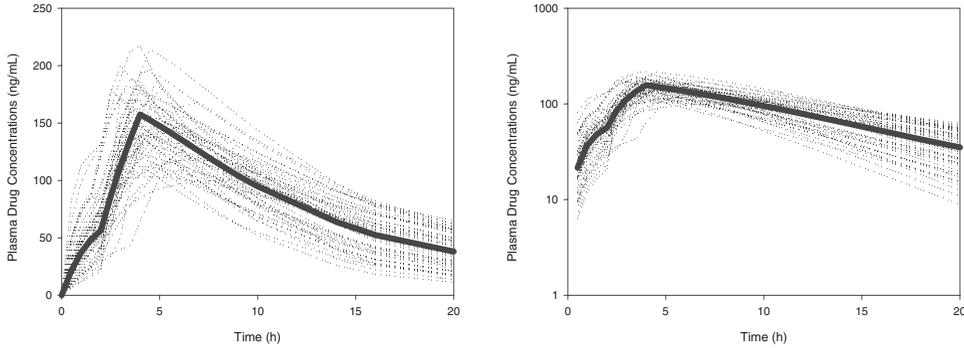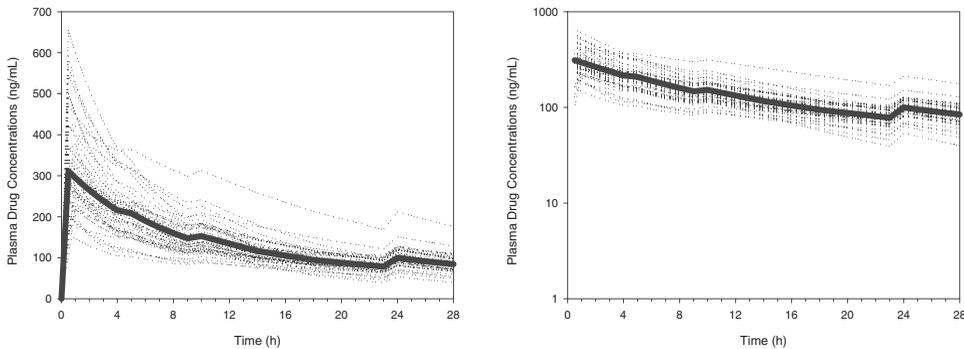
FIGURE 13.5   Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming a mixture of first-order and zero-order absorption types. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.

$$\frac{dA(1)}{dt} = -WB \cdot A(1) \tag{13.16}$$

$$\frac{dA(2)}{dt} = WB \cdot A(1) - \frac{CL}{V}A(2) \tag{13.17}$$

with

$$WB = 1 - fe^{\left(-(k_{a1} \cdot TIME)^{\gamma_1}\right)} - (1 - f)e^{\left(-(k_{a2} \cdot TIME)^{\gamma_2}\right)} \tag{13.18}$$

with the initial conditions at time zero

$$A(1) = DOSE$$

$$A(2) = 0$$

and where k_{a1} is the first-order absorption constant rate for the first phase, γ_1 is the shape factor for the first phase, k_{a2} is the first-order absorption constant rate for the second phase, γ_2 is the shape factor for the second phase, f is the fraction of the dose in the first phase, and CL and V are as defined above.

The plasma drug profiles in 50 subjects, following oral administration of the drug X and assuming Weibull-type absorption, are plotted in Figure 13.6 (one Weibull function) and in Figure 13.7 (two Weibull functions). Corresponding NONMEM control files and data sets are in Appendix 13.5.

12.3.6   Other Atypical Absorption Models

The models presented above will allow the adequate description of most drug absorption profiles; however, on occasion more complex models are needed. These would include the saturable time-constraint absorption model with a storage compartment (67), an extended compartmental absorption and transit model for

FIGURE 13.6   Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming a Weibull-type (one function) absorption. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.
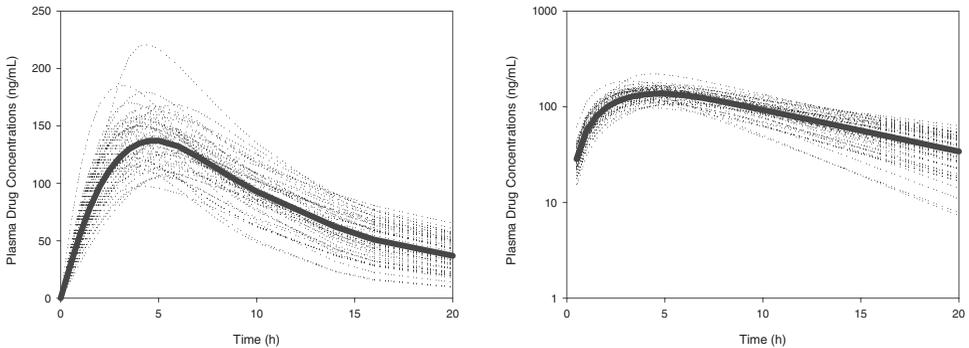
FIGURE 13.7   Plot of simulated plasma concentrations of drug X in 50 subjects, following a single oral dose, assuming a Weibull-type (two functions) absorption. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population predictions for a typical subject.

saturable small intestinal absorption (68), or an Erlang frequency distribution, which describes asymmetric s-shape absorption profiles (69), inverse Gaussian input (70, 71), and gamma distribution (72). Dual site absorption models can be described by models involving the description of two processes (two parallel first-order, mixed zero-order and first-order, etc.).

13.4   ENTEROHEPATIC RECYCLING MODEL

Enterohepatic recycling occurs when there is biliary excretion followed by intestinal reabsorption of a compound, sometimes with hepatic conjugation and intestinal deconjugation. Multiple peaks in a plasma concentration–time profile may be a consequence of this recycling. Determining whether the multiple peaks are due

to irregular absorption or enterohepatic recycling is of crucial importance, since the latter may significantly affect the half-life (prolongation), the exposure, and the bioavailability of the drug (73–75). Several methods exist for assessing the existence and degree of the enterohepatic recycling. Usually, when comparing the plasma concentration–time profiles obtained after oral and intravenous administration, if enterohepatic recycling is present the multiple peaks should be observed for both routes of administration. When intravenous administration is problematic, the use of charcoal administration after oral administration of the drug (76–78) or use of biliary tube drainage (79–81) may help in determining the relevance and the extent of the enterohepatic recycling phenomenon.

A general treatment of enterohepatic recycling based on the fraction of the drug in systemic circulation that is excreted in the bile and the fraction of drug reabsorbed from the gut that reaches systemic circulation in each enterohepatic cycle has been proposed (82). However, the description of enterohepatic recycling is often done through compartmental models. The recycling models can be divided between models with gallbladder emptying at regular intervals and models with gallbladder emptying at irregular intervals (83, 84). Irregular biliary emptying models are mathematically complex because the onset of gallbladder emptying needs to be known. Hence, these models are often limited to one or two recirculations (48, 85–88). The latter models are related to physiology since gallbladder emptying starts when food enters the region of the upper gastrointestinal tract (89). Moreover, enterohepatic recycling is often assessed following administration of a single dose but less often after multiple dosing (84).

Recent population modeling papers described multiple peaks due to enterohepatic recycling (85, 89). We present a model that takes into account multiple peaks at irregular intervals, after a single oral dose of a drug Y. Data from a Phase 1 study were analyzed. Meal times were known for the first 24 hours (4, 9, and 23 hours postdose). Drug Y pharmacokinetics was described by a zero-order absorption process and a two-compartment disposition model with first-order elimination. The drug was assumed to accumulate in the gallbladder following a first-order rate constant. Gallbladder emptying was postulated to be at meal times (4, 9, and 23 hours postdose) directly into the central compartment with zero-order kinetics (Figure 13.8).
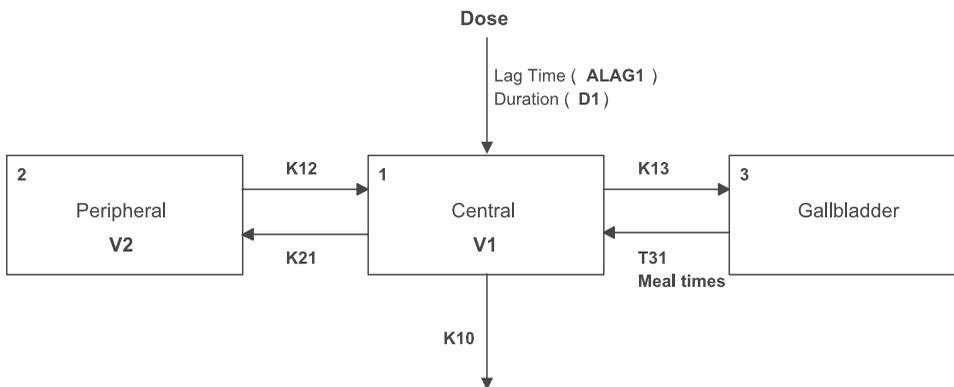
FIGURE 13.8 Enterohepatic recycling model.

The system of differential equations describing the amounts of the drug in the different compartments is as follows.

For $ALAG1 <$ time $\leq D1$, and outside the gallbladder emptying period,

$$\frac{dA(1)}{dt} = \frac{DOSE}{D1} - \frac{Q2}{V2}A(2) - \left[\frac{CL}{V1} + \frac{Q2}{V1} + K_{13}\right]A(1) \tag{13.19}$$

$$\frac{dA(2)}{dt} = \frac{Q2}{V1}A(1) - \frac{Q2}{V2}A(2) \tag{13.20}$$

$$\frac{dA(3)}{dt} = K_{13}A(1) \tag{13.21}$$

For time $> D1$, and outside the gallbladder emptying period,

$$\frac{dA(1)}{dt} = -\frac{Q2}{V2}A(2) - \left[\frac{CL}{V1} + \frac{Q2}{V1} + K_{13}\right]A(1) \tag{13.22}$$

$$\frac{dA(2)}{dt} = \frac{Q2}{V1}A(1) - \frac{Q2}{V2}A(2) \tag{13.23}$$

$$\frac{dA(3)}{dt} = K_{13}A(1) \tag{13.24}$$

For $4 <$ time $\leq (4 + T_{31})$, $9 <$ time $\leq (9 + T_{31})$, $23 <$ time $\leq (23 + T_{31})$,

$$\frac{dA(1)}{dt} = -\frac{Q2}{V2}A(2) - \left[\frac{CL}{V1} + \frac{Q2}{V1} + K_{13}\right]A(1) + \frac{A(3)}{T_{31}} \tag{13.25}$$

$$\frac{dA(2)}{dt} = \frac{Q2}{V1}A(1) - \frac{Q2}{V2}A(2) \tag{13.26}$$

$$\frac{dA(3)}{dt} = K_{13}A(1) - \frac{A(3)}{T_{31}} \tag{13.27}$$

with the initial condition at time zero

$$A(1) = 0$$
$$A(2) = 0$$
$$A(3) = 0$$

and where $D1$ is the duration of the zero-order absorption, $ALAG1$ is the lag time, K_{13} is the first-order rate constant for drug accumulation in gallbladder, T_{31} is the duration of the zero-order emptying process of the gallbladder into the central compartment, and CL, $Q2$, $V1$, and $V2$ are disposition and elimination parameters.

**TABLE 13.3   Population Parameters of Drug Y, Following a Single Oral
Dose of 10 mg, Obtained After Analyzing a Data Pool of Phase 1 Study
(Residual Variability Not Taken into Account)**

Parameters	Estimates	Between-Subject Error
Disposition parameters		
CL (L/h)	0.518	55.9% CV
$V1$ (L)	31.1	32.1% CV
$Q2$ (L/h)	2.23	76.4% CV
$V2$ (L)	47.1	—
Absorption parameters		
$D1$ (h)	0.196	70.7% CV
$ALAG1$ (h)	0.123	—
Enterohepatic recycling parameters		
K_{13} (h^{-1})	0.0236	—
T_{13} (h)	1.55	—

FIGURE 13.9   Plot of simulated plasma concentrations of drug Y in 50 subjects, following
a single oral dose, assuming zero-order absorption, and an enterohepatic recycling phenom-
enon. Normal (*left panel*) and semilog (*right panel*) scale. Thick line represents population
predictions for a typical subject.

Population parameter estimates with between-subject variability, obtained
from the analysis of the Phase 1 data, are presented in Table 13.3. Simulated
plasma drug profiles in 50 subjects, following a single 10 mg oral administration, are
plotted in Figure 13.9. Corresponding NONMEM control file and data set are in
Appendix 13.6.

13.5   SUMMARY

A wide array of drug absorption models and powerful computational facilities
are available to population and traditional PK modelers, yet adequate absorption
models are seldom available, presented in the literature, or estimated during drug
development. The factor limiting the development of adequate absorption models is

the design and execution of studies that will allow precise characterization of drug absorption. Given the importance of characterizing absorption, more effort should be expended on developing these models in the future.

The need for developing population models of absorption and/or enterohepatic recycling should be carefully assessed, keeping in mind the physiology principles behind the absorption process and the objectives of the analysis. Models can be developed to describe and summarize the data. When one wishes to use a model in order to simulate data not yet observed, the model must be completed by exploring and quantifying the variability (90–92).

The population approach for the description of the absorption of drugs has several advantages. First, the modeling approach is an objective assessment (statistical chi-square test) for the quantification of the factors affecting the absorption of a compound, especially when data come from different sources (pooled data from several studies). Second, with the use of the Bayesian estimation, it is possible to get individual PK parameters and then perform PK/PD analyses.

For clinical purposes, the first step is to classify factors influencing drug absorption with respect to the alterations in time course and magnitude of plasma concentrations with normal (healthy subjects vs. patients) or controlled conditions (food effect, drug–drug interactions). Historically, the characterization of drug absorption has been described empirically and lacked physiological rationale. In the future more attention should be paid to atypical absorption and improved study designs should be executed that better characterize the absorption profile.

REFERENCES

1. R. A. Prentis, Y. Lis, and S. R. Walker, Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). *Br J Clin Pharmacol* **25**:387–396 (1988).

2. R. T. Borchardt, P. L. Smith, and G. Wilson, General principles in the characterization and use of model systems for biopharmaceuticals studies. *Pharm Biotechnol* **8**:1–11 (1996).

3. L. J. Lesko, M. Rowland, C. C. Peck, and T. F. Blaschke, Optimizing the science of drug development: opportunities for better candidates selection and accelerated evaluation in humans. *J Clin Pharmacol* **40**:803–814 (2000).

4. C. C. Peck, W. H. Barr, L. Z. Benet, J. Collins, R. E. Desjardins, D. E. Furst, J. G. Harter, G. Levy, T. Ludden, and J. H. Rodman, Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in drug development. *J Clin Pharmacol* **34**:111–119 (1994).

5. C. A. Lipinsky, Drug-like properties and the cause of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* **44**:235–249 (2000).

6. B. J. Aungst, Novel formulation strategies for improving oral bioavailability of drugs with poor membrane permeation or presystemic metabolism. *J Pharm Sci* **82**:979–987 (1993).

7. H. Zhou, Pharmacokinetic strategies in deciphering atypical drug absorption profiles. *J Clin Pharmacol* **43**:211–227 (2003).

8. H. Zhou, Analysis of absorption kinetic data, in *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis*, Vol. 1, P. L. Bonate and D. R. Howard (Eds.). AAPS Press, New York, 2004, pp. 383–422.

9. W. L. Chiou, The rate and extent of oral bioavailability versus the rate and extent of oral absorption: clarification and recommendation of terminology. *J Pharmacokinet Pharmacodyn* **28**:3–6 (2001).

10. C. Duquesnoy, L. F. Lacey, O. N. Keene, and A. Bye, Evaluation of different partial AUCs as indirect measures of rate of drug absorption in comparative pharmacokinetic studies. *Eur J Pharm Sci* **6**:259–264 (1998).

11. D. Fleisher, C. Li, Y. Zhou, L. H. Pao, and A. Karim, Drug, meal and formulation interactions influencing drug absorption after oral administration clinical implications. *Clin Pharmacokinet* **36**:233–254 (1999).

12. C. H. Kleinbloesem, P. Van Brummelen, M. Danhof, et al., Rate of increase in the plasma concentration of nifedipine as a major determinant of its hemodynamics effect in humans. *Clin Pharm Ther* **41**:26–30 (1987).

13. E. H. Cox, C. Veyrat-Follet, S. L. Beal, E. Fuseau, S. Kenkare, and L. B. Sheiner, A population pharmacokinetic–pharmacodynamic analysis of repeated measures time-to-event pharmacodynamic responses: the antiemetic effect of ondansetron. *J Pharmacokinet Biopharm* **27**:625–644 (1999).

14. S. L. Collins, C. C. Faura, R. A. Moore, and H. J. McQuay, Peak plasma concentration after oral morphine: systemic review. *J Pain Symtom Manage* **16**:388–402 (1998).

15. E. Fuseau, O. Petricoul, A. Sabin, A. Pereira, S. O'Quinn, S. Thein, M. Leibowitz, H. Purdon, S. McNeal, R. Salonen, A. Metz, and P. Coates, Effect of encapsulation on absorption of sumatriptan tablets: data from healthy volunteers and patients during a migraine. *Clin Ther* **23**:242–251 (2001).

16. D. A. Sica, Pharmacotherapy in congestive heart failure: drug absorption in the management of congestive heart failure: loop diuretics. *Congestive Heart Failure* **9**:287–292 (2003).

17. L. K. M. De Bruyne, Mechanisms and management of diuretic resistance in congestive heart failure. *Postgrad Med J* **79**:268–271 (2003).

18. E. Rubinstein and J. Camm, Cardiotoxicity of fluoroquinolones. *J Antimicrob Chemother* **49**:593–596 (2002).

19. J. L. Demolis, D. Kubitza, L. Tenneze, and C. Funck-Brentano, Effect of a single oral dose of moxifloxacin (400 mg and 800 mg) on ventricular repolarization in healthy subjects. *Clin Pharmacol Ther* **68**:658–666 (2000).

20. S. Barriere, F. Genter, E. Spencer, M. Kitt, D. Hoelscher, and J. Morganroth, Effects of a new antibacterial, telavancin, on cardiac repolarization (QTc interval duration) in healthy subjects. *J Clin Pharmacol* **44**:689–695 (2004).

21. C. MacDougall and B. J. Guglielmo, Pharmacokinetics of valaciclovir. *J Antimicrob Chemother* **53**:899–901 (2004).

22. A. Schwarz and A. Perez-Canto, Nephrotoxicity of antiinfective drugs. *Int J Clin Pharmacol Ther* **36**:164–167 (1998).

23. H. Echizen and M. Eichelbaum, Clinical pharmacokinetics of verapamil, nifedipine and diltiazem. *Clin Pharmacokinet* **11**:425–449 (1986).

24. V. Luckow and O. Della Paschoa, PKPD modeling of high-dose ditliazem—absporption-rate dependency of the hysteresis loop. *Int J Clin Pharmacol Ther* **35**:418–425 (1997).

25. M. D. Cheitlin, A. M. Hutter, R. G. Brindis, P. Ganz, S. Kaul, R. O. Russell, and R. M. Zusman, Use of sildenafil (Viagra) in patients with cardiovascular disease. *Circulation* **99**:168–177 (1999).

26. D. J. Nichols, G. J. Muirhead, and J. A. Harness, Pharmacokinetics of sildenafil after single oral doses in healthy male subjects: absolute bioavailability, food effects and dose proportionality. *Br J Clin Pharmacol* **53**(Suppl 1):5S–12S (2002).

27. M. F. Delgado Iribarnegaray, D. Santo Bueldga, M. J. Garcia Sanchez, M. J. Otero, A. C. Falcao, and A. Dominguez-Gil, Carbamazepine population pharmacokinetics in children: mixed-effect models. *Ther Drug Monit* **19**:132–139 (1997).

28. D. S. Stein, Y. Lou, M. Johnson, S. Randall, and S. Blanche, Pharmacokinetic and pharmacodynamic analysis of amprenavir-containing combination therapy in HIV-1-infected children. *J Clin Pharmacol* **44**:1301–1308 (2004).

29. D. Hickman, S. Vasavanonda, G. Nequist, L. Colletti, W. M. Kati, R. Bertz, A. Hsu, and D. J. Kempf, Estimation of serum-free 50-percent inhibitory concentrations for human immunodeficiency virus protease inhibitors lopinavir and ritonavir. *Antimicrob Agents Chemother* **48**:2911–2917 (2004).

30. E. Iisalo, Clinical pharmacokinetics of digoxin. *Clin Pharmacokinet* **2**:1–16 (1977).

31. M. Weiss and W. Kang, Inotropic effect of digoxin in humans: mechanistics pharmacokinetic/pharmacodynamic model based on slow receptor binding. *Pharm Res* **21**:231–236 (2004).

32. G. E. Granero, C. Ramachanda, and G. L. Amidon, in *Drug Bioavailability. Estimation of Solubility, Permeability, Absorption, and Bioavailability*, Vol. 18, H. Van de Waterbeemd, H. Lennernäs, and P. Artursson (Eds.). Wiley–VCH, Hoboken, NJ, 2004, pp. 191–214.

33. H. Van de Waterbeemd, H. Lennernäs, and P. Artursson (Eds.), *Drug Bioavailability. Estimation of Solubility, Permeability, Absorption, and Bioavailability*, Vol. 18. Wiley–VCH, Hoboken, NJ, 2004, pp. 3–20.

34. M. N. Martinez and G. L. Amidon, A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. *J Clin Pharmacol* **42**:620–643 (2002).

35. G. L. Amidon, H. Lennernas, V. P. Shah, and J. R. Crison, Theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res* **12**:413–420 (1995).

36. K. C. Johnson and A. C. Swindell, Guidance in the setting of drug particle size specifications to minimize variability in absorption. *Pharm Res* **13**:1795–1798 (1996).

37. B. Agoram, W. S. Woltosz, and M. B. Bolger, Predicting the impact of physiological and biochemical processes on oral drug bioavailability. *Adv Drug Deliv Rev* **50**(Suppl 1):S41–S67 (2001).

38. P. Stenberg, C. A. Bergstrom, K. Luthman, and P. Artursson, Theoretical predictions of drug absorption in drug discovery and development. *Clin Pharmacokinet* **41**:877–899 (2002).

39. N. M. Idkaidek, G. L. Amidon, D. E. Smith, N. M. Najib, and M. M. Hassan, Determination of the population pharmacokinetic parameters of sustained-release and enteric-coated oral formulations, and the suppository formulation of diclofenac sodium by simultaneous data fitting using NONMEM. *Biopharm Drug Dispos* **19**:169–174 (1998).

40. C. Csajka, T. Buclin, K. Fattinger, H. R. Brunner, and J. Biollaz, Population pharmacokinetic–pharmacodynamic modeling of angiotensin receptor blockade in healthy volunteers. *Clin Pharmacokinet* **41**:137–152 (2002).

41. Y. M. Wang and R. H. Reuning, An experimental design strategy for quantitating complex pharmacokinetic models: enterohepatic circulation with time-varying gallbladder emptying as an example. *Pharm Res* **9**:169–177 (1992).

42. Y. Plusquellec, G. Campiston, S. Staveris, J. Barre, L. Jung, J. P. Tillement, and G. Houin, A double-peak phenomenon in the pharmacokinetics of veralipride after oral administration: a double-site model for drug absorption. *J Pharmacokinet Biopharm* **15**:225–239 (1987).

43. A. M. Holgate and N. M. Read, Effect of ileal infusion of intralipid on gastrointestinal transit, ileal flow rate, and carbohydrate absorption in humans after ingestion of a liquid meal. *Gastroenterology* **88**:1005–1011 (1985).

44. R. L. Oberle and G. L. Amidon, The influence of variable gastric emptying and intestinal rates on the plasma level curve of cimetidine: an explanation for the double peak phenomenon. *J Pharmacokinet Biopharm* **15**:529–544 (1987).

45. P. H. Marathe, E. P. Sandefer, G. E. Kollia, D. S. Greene, R. H. Barbhaiya, R. A. Lipper, R. C. Page, W. Doll, U. Y. Ryo, and G. A. Digenis, In vivo evaluation of the absorption and gastrointestinal transit of avitriptan in fed and fasted subjects using gamma scintigraphy. *J Pharmacokinet Biopharm* **26**:1–20 (1998).

46. E. Lipka, I. D. Lee, P. Langguth, H. Spahn-Langguth, E. Mutschler, and G. L. Amidon, Celiprolol double-peak and gastric motility: non-linear mixed effect modeling of bioavailability data obtained in dogs. *J Pharmacokinet Biopharm* **23**:267–287 (1995).

47. W. N. Charman, M. Rogge, A. W. Body, W. H. Barr, and B. M. Berger, Absorption of danazol after administration to different sites of the gastrointestinal tract and the relationship to single- and double-peak phenomena in the plasma profiles. *J Clin Pharmacol* **33**:1207–1213 (1993).

48. Y. Wang, A. Roy, L. Sun, and C. E. Lau, A double-peak phenomenon in the pharmacokinetics of alprazolam after oral administration. *Drug Metab Dispos* **27**:855–859 (1999).

49. P. Veng Pedersen, Pharmacokinetic analysis by linear system approach I: cimetidine bioavailability and second peak phenomenon. *J Pharm Sci* **70**:32–38 (1981).

50. I. F. Troconiz, S. Armenteros, M. V. Planelles, J. Benitez, R. Calvo, and R. Dominguez, Pharmacokinetic–pharmacodynamic modeling of the antipyretic effect of two oral formulations of ibuprofen. *Clin Pharmacokinet* **38**:505–518 (2000).

51. K. Murata, K. Noda, K. Kohno, and M. Samejima, Pharmacokinetic analysis of concentration data of drugs with irregular absorption profiles using multi-fraction absorption models. *J Pharm Sci* **76**:109–113 (1987).

52. K. Murata, H. Yamahara, M. Kobayashi, K. Noda, and M. Samejima, Pharmacokinetics of an oral sustained-release diltiazem preparation. *J Pharm Sci* **78**:960–963 (1989).

53. I. Mahmood, Pharmacokinetic analysis of the absorption characteristics of diclofenac sodium in man by use of a multi-segment absorption model. *J Pharm Pharmacol* **48**:1260–1263 (1996).

54. T. Funaki, N. Watari, S. Furata, and N. Kaneniwa, Analysis of the absorption characteristics of cimetidine with the use of the multi-segment absorption model. *Int J Pharm* **43**:59–65 (1988).

55. K. Murata and K. Noda, Pharmacokinetic analysis of an oral sustained-release diltiazem preparation using multifraction absorption models. *Pharm Res* **10**:757–762 (1993).

56. V. F. Cosson and E. Fuseau, Mixed effect modeling of sumatriptan pharmacokinetics during drug development: II. From healthy subjects to phase 2 dose ranging in patients. *J Pharmacokinet Biopharm* **27**:149–171 (1999).

57. F. Bouzom, C. Laveille, H. Merdjan, and R. Jochemsen, Use of nonlinear mixed effect modeling for the meta-analysis of preclinical pharmacokinetic data: application to S 20342 in the rat. *J Pharm Sci* **89**:603–613 (2000).

58. T. M. Garrigues, U. Martin, J. E. Peris-Ribera, and L. F. Prescott, Dose-dependent absorption and elimination of cefadroxil in man. *Eur J Clin Pharmacol* **41**:179–183 (1991).

59. N. H. Holford, R. J. Ambros, and K. Stoeckel, Models for describing absorption rate and estimating extent of bioavailability: application to cefetamet pivoxil. *J Pharmacokinet Biopharm* **20**:421–442 (1992).

60. S. Rietbrock, P. G. Merz, U. Fuhr, S. Harder, J. P. Marschner, D. Loew, and J. Biehl, Absorption behavior of sulpiride described using Weibull functions. *Int J Clin Pharmacol Ther* **33**:299–303 (1995).

61. D. Hartmann, D. Gysel, U. C. Dubach, and I. Forgo, Pharmacokinetic modeling of the plasma concentration–time profile of the vitamin retinyl palmitate following intramuscular administration. *Biopharm Drug Dispos* **11**:689–700 (1990).

62. D. Smith, R. Enever, M. Dey, D. Latta, and R. Weierstall, Pharmacokinetics and bioavailability of medroxyprogesterone acetate in the dog and the rat. *Biopharm Drug Dispos* **14**:341–355 (1993).

63. F. Bressolle, R. Gomeni, R. Alric, M. J. Royer-Morrot, and J. Necciari, A double Weibull input function describes the complex absorption of sustained-release oral sodium valproate. *J Pharm Sci* **83**:1461–1464 (1994).

64. S. Appel-Dingemanse, M. O. Lemarechal, A. Kumle, M. Hubert, and E. Legangneux, Integrated modeling of the clinical pharmacokinetics of SDZ HTF 919, a novel selective 5-HT4 receptor agonist, following oral and intravenous administration. *Br J Clin Pharmacol* **47**:483–491 (1999).

65. H. Zhou, S. Khalilieh, H. Lau, M. Guerret, S. Osborne, L. Alladina, A. L. Laurent, and J. F. McLeod, Effect of meal timing not critical for the pharmacokinetics of tegaserod (HTF 919). *J Clin Pharmacol* **39**:911–919 (1999).

66. C. Padoin, M. Tod, N. Brion, K. Louchahi, V. Le Gros, and O. Petitjean, Pharmacokinetics of amoxicillin coadministered with a saline–polyethylene glycol solution in healthy volunteers. *Biopharm Drug Dispos* **16**:169–176 (1995).

67. V. K. Piotrovskij, G. Paintaud, G. Alvan, and T. Trnovec, Modeling of the saturable time-constrained amoxicillin absorption in humans. *Pharm Res* **11**:1346–1351 (1994).

68. L. X. Yu and G. L. Amidon, Saturable small intestinal drug absorption in humans: modeling and interpretation of cefatrizine data. *Eur J Pharm Biopharm* **45**:199–203 (1998).

69. A. Rousseau, F. Leger, Y. Le Meur, F. Saint-Marcoux, G. Paintaud, M. Buchler, and P. Marquet, Population pharmacokinetic modeling of oral cyclosporin using NONMEM: comparison of absorption pharmacokinetic models and design of a Bayesian estimator. *Ther Drug Monit* **26**:23–30 (2004).

70. M. Weiss, A novel extravascular input function for the assessment of drug absorption in bioavailability studies. *Pharm Res* **13**:1457–1553 (1996).

71. X. Zhang, K. Nieforth, J. M. Lang, R. Rouzier-Panis, J. Reynes, A. Dorr, S. Kolis, M. R. Stiles, T. Kinchelow, and I. H. Patel, Pharmacokinetics of plasma enfuvirtide after subcutaneous administration to patients with human immunodeficiency virus: inverse Gaussian density absorption and 2-compartment disposition. *Clin Pharmacol Ther* **72**:10–19 (2002).

72. J. Debord, E. Risco, M. Harel, Y. Le Meur, M. Buchler, G. Lachatre, C. Le Guellec, and P. Marquet, Application of a gamma model of absorption to oral cyclosporin. *Clin Pharmacokinet* **40**:375–382 (2001).

73. M. S. Roberts, B. M. Magnusson, F. J. Burczynski, and M. Weiss, Enterohepatic circulation: physiological, pharmacokinetic and clinical implications. *Clin Pharmacokinet* **41**:751–790 (2002).

74. T. B. Vree and A. J. Van der Ven, Clinical consequences of the biphasic elimination kinetics for the diuretic effect of furosemide and its acyl glucuronide in humans. *J Pharm Pharmacol* **51**:239–248 (1999).

75. R. E. Bullingham, A. J. Nicholls, and B. R. Kamm, Clinical pharmacokinetics of mycophenolate mofetil. *Clin Pharmacokinet* **34**:429–455 (1998).

76. K. Elomaa, S. Ranta, J. Tuominen, and P. Lahteenmaki, The possible role of entero-hepatic cycling on bioavailability of norethisterone and gestodene in women using combined oral contraceptives. *Contraception* **63**:13–18 (2001).

77. Y. J. Sue and M. Shannon, Pharmacokinetics of drugs in overdose. *Clin Pharmacokinet* **23**:93–105 (1992).

78. D. J. Back and M. L. Orme, Pharmacokinetic drug interactions with oral contraceptives. *Clin Pharmacokinet* **18**:472–484 (1990).

79. W. A. Colburn, F. M. Vane, C. J. Bugge, D. E. Carter, R. Bressler, and C. W. Ehmann, Pharmacokinetics of ^{14}C-isotretinoin in healthy volunteers and volunteers with biliary T-tube drainage. *Drug Metab Dispos* **13**:327–332 (1985).

80. A. Hellstern, D. Hellenbrecht, R. Saller, M. Gatzen, G. Achtert, P. Brockmann, and H. J. Hausleiter, Minimal biliary excretion and enterohepatic recirculation of meto-clopramide in patients with extrahepatic cholestasis. *Eur J Clin Pharmacol* **45**:415–418 (1993).

81. R. Lenzen, A. Bahr, H. Eichstadt, U. Marshall, W. O. Bechstein, and P. Neuhaus, In liver transplantation T tube bile represents total bile flow: physiological, and scintigraphic studies on biliary secretion of organic anions. *Liver Transpl Surg* **5**:8–15 (1999).

82. J. E. Peris-Ribera, F. Torres-Molina, M. C. Garcia-Carbonell, J. C. Aristorena, and L. Granero, General treatment of the enterohepatic recirculation of drugs and its influence on the area under the plasma level curves, bioavailability, and clearance. *Pharm Res* **9**:1306–1313 (1992).

83. R. L. Semmes and D. D. Shen, A reversible clearance model for the enterohepatic circulation of drug and conjugate metabolite pair. *Drug Metab Dispos* **18**:80–87 (1990).

84. T. Wajima, Y. Yano, and T. Oguma, A pharmacokinetic model for analysis of drug disposition profiles undergoing enterohepatic circulation. *J Pharm Pharmacol* **54**:929–934 (2002).

85. T. Funaki, Enterohepatic circulation model for population pharmacokinetic analysis. *J Pharm Pharmacol* **51**:1143–1148 (1999).

86. B. Ploeger, T. Mensinga, A. Sips, J. Meulenbelt, and J. DeJongh, A human physiologically-based model for glycyrrhzic acid, a compound subject to presystemic metabolism and enterohepatic cycling. *Pharm Res* **17**:1516–1525 (2000).

87. K. Strandgarden, P. Hoglund, L. Gronquist, L. Svensson, and P. O. Gunnarsson, Absorption and disposition including enterohepatic circulation of (^{14}C) roquinimex after oral administration to healthy volunteers. *Biopharm Drug Dispos* **21**:53–67 (2000).

88. Y. Plusquellec, R. Arnaud, S. Saivin, T. A. Shepard, I. Carrie, P. Hermann, J. Souhait, and G. Houin, Enterohepatic recirculation of the new antihypertensive drug UP 269-6 in humans. A possible model to account for multiple plasma peaks. *Arzneimittelforschung* **48**:138–144 (1998).

89. F. Ezzet, G. Krishna, D. B. Wexler, P. Statkevich, T. Kosoglou, and V. K. Batra, A population pharmacokinetic model that describes multiple peaks due to enterohepatic recirculation of ezetimibe. *Clin Ther* **23**:871–885 (2001).

90. N. H. Holford, H. C. Kimko, J. P. Monteleone, and C. C. Peck, Simulation of clinical trials. *Annu Rev Pharmacol Toxicol* **40**:209–234 (2000).

91. K. F. Rooney, E. Snoeck, and P. H. Watson, Modeling and simulation in clinical drug development. *Drug Discov Today* **6**:802–806 (2001).

92. M. Sale, Modeling and simulation in drug development, promise and reality. *Drug Discov World* **2**:47–50 (2001).

APPENDIX 13.1   FIRST-ORDER ABSORPTION

NONMEM Input

```
$PROB First-Order Absorption
$INPUT ID TIME DV AMT CMT EVID MDV
$DATA EX1_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=2
COMP=(DEPOT,DEFDOSE)  COMP=(CENTRAL,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

KA = THETA(3)*EXP(ETA(3))

K12 = KA
K20 = CL/V

S2=V/1000
$DES

DADT(1) = -K12*A(1)
DADT(2) = K12*A(1) - K20*A(2)

$ERROR
;additive error
 Y=F+ERR(1)
 IPRED=F

$THETA
5 ;1 CL
50 ;2 V2

0.5 ;3 KA

$OMEGA

0.04 ; 20% CV for CL
0.04 ; 20% CV for V
0.36 ; 60% CV for KA
```

```
$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM

$TABLE NOPRINT ONEHEADER FILE=FIRST.PAR

ID TIME DV AMT EVID MDV KA CL V IPRED
```

Data Set

#ID	TIME	DV	AMT	CMT	EVID	MDV
1	0	.	10	1	1	1
1	0.5	.	.	2	0	0
1	1	.	.	2	0	0
1	1.5	.	.	2	0	0
1	2	.	.	2	0	0
1	2.5	.	.	2	0	0
1	3	.	.	2	0	0
1	3.5	.	.	2	0	0
1	4	.	.	2	0	0
1	4.5	.	.	2	0	0
1	5	.	.	2	0	0
1	6	.	.	2	0	0
1	7	.	.	2	0	0
1	8	.	.	2	0	0
1	9	.	.	2	0	0
1	10	.	.	2	0	0
1	14	.	.	2	0	0
1	16	.	.	2	0	0
1	24	.	.	2	0	0
2	0	.	10	1	1	1
2	0.5	.	.	2	0	0
2	1	.	.	2	0	0

APPENDIX 13.2   ZERO-ORDER ABSORPTION

NONMEM Input

```
$PROB Zero-Order Absorption
$INPUT ID TIME DV AMT CMT RATE EVID MDV
$DATA EX2_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=1
COMP=(DEPOT,DEFDOSE,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

;duration (h) of the zero-order process
D1 = THETA(3)*EXP(ETA(3))

K10 = CL/V

S1=V/1000

$DES

DADT(1) = - K10*A(1)

$ERROR
;additive error
Y=F+ERR(1)
IPRED=F

$THETA
5 ;1 CL
50 ;2 V

2 ;3 D1

$OMEGA

  0.04 ; 20% CV for CL
  0.04 ; 20% CV for V
  0.36 ; 60% CV for D1

$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM

$TABLE NOPRINT ONEHEADER FILE=ZERO.PAR
ID TIME DV AMT EVID MDV D1 CL V IPRED
```

Data Set

#ID	TIME	DV	AMT	CMT	RATE	EVID	MDV
1	0	.	10	1	–2	1	1
1	0.5	.	.	1		0	0
1	1	.	.	1		0	0
1	1.5	.	.	1		0	0
1	2	.	.	1		0	0
1	2.5	.	.	1		0	0
1	3	.	.	1		0	0
1	3.5	.	.	1		0	0
1	4	.	.	1		0	0
1	4.5	.	.	1		0	0
1	5	.	.	1		0	0
1	6	.	.	1		0	0
1	7	.	.	1		0	0
1	8	.	.	1		0	0
1	9	.	.	1		0	0
1	10	.	.	1		0	0
1	14	.	.	1		0	0
1	16	.	.	1		0	0
1	24	.	.	1		0	0
2	0	.	10	1	–2	1	1
2	0.5	.	.	1		0	0
2	1	.	.	1		0	0

APPENDIX 13.3 TWO PARALLEL FIRST-ORDER ABSORPTIONS

NONMEM Input

```
$PROB Two Parallel first-order Absorption
$INPUT ID TIME DV AMT CMT EVID MDV
$DATA EX3_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=3
COMP=(DEPOT1,DEFDOSE) COMP=(DEPOT2) COMP=(CENTRAL,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

KA1 = THETA(3)*EXP(ETA(3))

KA2 = THETA(4)*EXP(ETA(4))

;lag time for second process
ALAG2 = THETA(5)*EXP(ETA(5))

;F1 fraction of the dose absorbed through the first process
F1 = THETA(6)*EXP(ETA(6))

;F2 fraction of the dose absorbed through the second process
F2 = 1-F1

K13 = KA1
K23 = KA2
K30 = CL/V

S3=V/1000

$DES

DADT(1) = -K13*A(1)
DADT(2) = -K23*A(2)
DADT(3) = K13*A(1) + K23*A(2) - K30*A(3)

$ERROR
;additive error
Y=F+ERR(1)
IPRED=F

$THETA
5 ;1 CL
50 ;2 V
```

```
0.8 ;3 KA1
0.6 ;4 KA2
5 ;5 ALAG2
0.5 ;6 F1

$OMEGA

  0.04 ; 20% CV for CL
  0.04 ; 20% CV for V
  0.36 ; 60% CV for KA1
  0.36 ; 60% CV for KA2
  0.04 ; 20% CV for ALAG2
  0.04 ; 20% CV for F1

$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM

$TABLE NOPRINT ONEHEADER FILE=PARAL.PAR
ID TIME DV AMT EVID MDV
KA1 KA2 ALAG2 F1 F2 CL V IPRED
```

Data Set

#ID	TIME	DV	AMT	CMT	EVID	MDV
1	0	.	10	1	1	1
1	0	.	10	2	1	1
1	0.5	.	.	3	0	0
1	1	.	.	3	0	0
1	1.5	.	.	3	0	0
1	2	.	.	3	0	0
1	2.5	.	.	3	0	0
1	3	.	.	3	0	0
1	3.5	.	.	3	0	0
1	4	.	.	3	0	0
1	4.5	.	.	3	0	0
1	5	.	.	3	0	0
1	6	.	.	3	0	0
1	7	.	.	3	0	0
1	8	.	.	3	0	0
1	9	.	.	3	0	0
1	10	.	.	3	0	0
1	14	.	.	3	0	0
1	16	.	.	3	0	0
1	24	.	.	3	0	0
2	0	.	10	1	1	1
2	0	.	10	2	1	1
2	0.5	.	.	3	0	0
2	1	.	.	3	0	0

APPENDIX 13.4   MIXTURE OF FIRST-ORDER AND ZERO-ORDER ABSORPTION

NONMEM Input

```
$PROB Mix first-zero order Absorption
$INPUT ID TIME DV AMT CMT RATE EVID MDV
$DATA EX4_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=2
COMP=(DEPOT,DEFDOSE) COMP=(CENTRAL,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

KA = THETA(3)*EXP(ETA(3))
D2 = THETA(4)*EXP(ETA(4))

;ALAG2 lag time for zero-order process
ALAG2 = THETA(5)*EXP(ETA(5))

;F1 fraction of the dose absorbed through the first-order process
F1 = THETA(6)*EXP(ETA(6))

;F2 fraction of the dose absorbed through the zero-order process
F2 = 1-F1

K12 = KA
K20 = CL/V

S2=V/1000

$DES

DADT(1) = -K12*A(1)
DADT(2) = K12*A(1) - K20*A(2)

$ERROR
;additive error
 Y=F+ERR(1)
  IPRED=F

$THETA
5 ;1 CL
50 ;2 V2

0.5 ;3 KA
2 ;4 D2
2 ;5 ALAG2
```

```
0.5 ;6 F1

$OMEGA

 0.04 ; 20% CV for CL
 0.04 ; 20% CV for V
 0.36 ; 60% CV for KA
 0.36 ; 60% CV for D1
 0.04 ; 20% CV for ALAG2
 0.04 ; 20% CV for F1

$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM

$TABLE NOPRINT ONEHEADER FILE=MIX.PAR
 ID TIME DV AMT EVID MDV
 KA D2 ALAG2 F1 F2 CL V IPRED
```

Data Set

#ID	TIME	DV	AMT	CMT	RATE	EVID	MDV
1	0	.	10	1	.	1	1
1	0	.	10	2	–2	1	1
1	0.5	.	.	2		0	0
1	1	.	.	2		0	0
1	1.5	.	.	2		0	0
1	2	.	.	2		0	0
1	2.5	.	.	2		0	0
1	3	.	.	2		0	0
1	3.5	.	.	2		0	0
1	4	.	.	2		0	0
1	4.5	.	.	2		0	0
1	5	.	.	2		0	0
1	6	.	.	2		0	0
1	7	.	.	2		0	0
1	8	.	.	2		0	0
1	9	.	.	2		0	0
1	10	.	.	2		0	0
1	14	.	.	2		0	0
1	16	.	.	2		0	0
1	24	.	.	2		0	0
2	0	.	10	1	.	1	1
2	0	.	10	2	–2	1	1
2	0.5	.	.	2		0	0
2	1	.	.	2		0	0

APPENDIX 13.5   WEIBULL-TYPE ABSORPTION

NONMEM Input for One Weibull Function

```
$PROB Weibull-type Absorption (n=1)
$INPUT ID TIME DV AMT CMT EVID MDV
$DATA EX5_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=2
COMP=(DEPOT,DEFDOSE) COMP=(CENTRAL,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

;Weibull parameters
KA1 = THETA(3)*EXP(ETA(3))
GAMA1 = THETA(4)*EXP(ETA(4))

;Weibull function
WB = 1-EXP((-(KA1*TIME)**GAMA1))

K20 = CL/V

S2=V/1000

$DES

DADT(1) = -WB*A(1)
DADT(2) = WB*A(1) - K20*A(2)

$ERROR
;additive error
Y=F+ERR(1)
IPRED=F

$THETA
5 ;1 CL
50 ;2 V2

0.4 ;3 KA1
4 ;4 GAMMA1

$OMEGA
```

```
 0.04 ; 20% CV for CL
 0.04 ; 20% CV for V
 0.36 ; 60% CV for KA1
 0.04 ; 20% CV for GAMMA1

$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM
$TABLE NOPRINT ONEHEADER FILE=WEIBULL3.PAR
  ID TIME DV AMT EVID MDV
  KA1 GAMA1 WB CL V IPRED
```

Data Set for One Weibull Function

#ID	TIME	DV	AMT	CMT	EVID	MDV
1	0	.	10	1	1	1
1	0.5	.	.	2	0	0
1	1	.	.	2	0	0
1	1.5	.	.	2	0	0
1	2	.	.	2	0	0
1	2.5	.	.	2	0	0
1	3	.	.	2	0	0
1	3.5	.	.	2	0	0
1	4	.	.	2	0	0
1	4.5	.	.	2	0	0
1	5	.	.	2	0	0
1	6	.	.	2	0	0
1	7	.	.	2	0	0
1	8	.	.	2	0	0
1	9	.	.	2	0	0
1	10	.	.	2	0	0
1	14	.	.	2	0	0
1	16	.	.	2	0	0
1	24	.	.	2	0	0
2	0	.	10	1	1	1
2	0.5	.	.	2	0	0
2	1	.	.	2	0	0

NONMEM Input for Two Weibull Functions

```
$PROB Weibull-type Absorption (n=2)
$INPUT ID TIME DV AMT CMT EVID MDV
$DATA EX5_DATA2.csv
IGNORE=#
```

```
$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=2
COMP=(DEPOT,DEFDOSE)  COMP=(CENTRAL,DEFOBS)

$PK

CL = THETA(1)*EXP(ETA(1))
V = THETA(2)*EXP(ETA(2))

;Weibull parameters
KA1 = THETA(3)*EXP(ETA(3))
GAMA1 = THETA(4)*EXP(ETA(4))

KA2 = THETA(5)*EXP(ETA(5))
GAMA2 = THETA(6)*EXP(ETA(6))

;FR1 Fraction of the dose absorbed through the first process
FR1 = THETA(7)*EXP(ETA(7))

;FR2 Fraction of the dose absorbed through the second process
FR2 = 1-FR1

;two Weibull functions
WB1 = EXP((-(KA1*TIME)**GAMA1))
WB2 = EXP((-(KA2*TIME)**GAMA2))
WB = 1-FR1*WB1 - FR2*WB2

K20 = CL/V

S2=V/1000

$DES

DADT(1) = -WB*A(1)
DADT(2) = WB*A(1) - K20*A(2)

$ERROR
;additive error
Y=F+ERR(1)
IPRED=F

$THETA
5 ;1 CL
50 ;2 V2

2.0 ;3 KA1
0.5 ;4 GAMMA1
0.2 ;5 KA2
4.0 ;6 GAMMA2
0.5 ;7 FR1

$OMEGA
```

```
0.04 ; 20% CV for CL
0.04 ; 20% CV for V
0.36 ; 60% CV for KA1
0.04 ; 20% CV for GAMMA1
0.36 ; 60% CV for KA2
0.04 ; 20% CV for GAMMA2
0.04 ; 20% CV for F1

$SIGMA 0.02

$SIMULATION (123456789) ONLYSIM
$TABLE NOPRINT ONEHEADER FILE=WEIBULL.PAR
 ID TIME DV AMT EVID MDV
 KA1 GAMA1 KA2 GAMA2 FR1 FR2 WB1 WB2 WB
 CL V IPRED
```

Data Set for Two Weibull Functions

#ID	TIME	DV	AMT	CMT	EVID	MDV
1	0	.	10	1	1	1
1	0	.	10	2	1	1
1	0.5	.	.	3	0	0
1	1	.	.	3	0	0
1	1.5	.	.	3	0	0
1	2	.	.	3	0	0
1	2.5	.	.	3	0	0
1	3	.	.	3	0	0
1	3.5	.	.	3	0	0
1	4	.	.	3	0	0
1	4.5	.	.	3	0	0
1	5	.	.	3	0	0
1	6	.	.	3	0	0
1	7	.	.	3	0	0
1	8	.	.	3	0	0
1	9	.	.	3	0	0
1	10	.	.	3	0	0
1	14	.	.	3	0	0
1	16	.	.	3	0	0
1	24	.	.	3	0	0
2	0	.	10	1	1	1
2	0	.	10	2	1	1
2	0.5	.	.	3	0	0
2	1	.	.	3	0	0

APPENDIX 13.6   ENTEROHEPATIC RECYCLING MODEL

NONMEM Input

```
$PROB Enterohepatic recycling
$INPUT ID TIME DV AMT CMT RATE EVID MDV
$DATA EHR_DATA.csv
IGNORE=#

$SUBROUTINE ADVAN6 TOL=4

$MODEL NCOMP=3
  COMP=(CENTRAL,DEFOBS,DEFDOSE)
  COMP=(PERIPH)
  COMP=(ACCUM)

$PK

;duration (h) of the zero-order process
D1=THETA(1)*EXP(ETA(1))

;lag time for the zero-order process
ALAG1=THETA(2)

;disposition parameters
CL=THETA(3)*EXP(ETA(2))
V1=THETA(4)*EXP(ETA(3))
Q2=THETA(5)*EXP(ETA(4))
V2=THETA(6)

;Enterohepatic recycling parameters
  K13=THETA(7)
  T31=THETA(8)

  S1=V1/1000

  K10=CL/V1
  K12=Q2/V1
  K21=Q2/V2

;times for release from the accumulation compartment
  MEA1=4+T31
  MEA2=9+T31
  MEA3=23+T31

$DES

; Default situation, only accumulation in the compt 3
  DADT(1) = K21*A(2)-(K10+K12+K13)*A(1)
  DADT(2) = K12*A(1)-K21*A(2)
  DADT(3) = K13*A(1)
```

```
; First release
  IF (TIME.GT.4.AND.TIME.LE.MEA1) THEN
  DADT(1) = K21*A(2)-(K10+K12+K13)*A(1)+A(3)/T31
  DADT(3) = K13*A(1)-A(3)/T31
  ENDIF

; Second release
  IF (TIME.GT.9.AND.TIME.LE.MEA2) THEN
    DADT(1) = K21*A(2)-(K10+K12+K13)*A(1)+A(3)/T31
    DADT(3) = K13*A(1)-A(3)/T31
  ENDIF

; Third release
  IF (TIME.GT.23.AND.TIME.LE.MEA3) THEN
    DADT(1) = K21*A(2)-(K10+K12+K13)*A(1)+A(3)/T31
  DADT(3) = K13*A(1)-A(3)/T31
  ENDIF

$ERROR

  A1=A(1)*1000
  A2=A(2)*1000
  A3=A(3)*1000

  IPRED=F
;combined error
Y=F*(1+ERR(1))+ERR(2)

$THETA

 0.196 ;1 D1
 0.123 ;2 Alag1
 0.518 ;3 CL
 31.1 ;4 V1
 2.23 ;5 Q2
 47.1 ;6 V2
 0.024 ;7 K13
 1.55 ;8 T31

$SIGMA
0.0402
0.220

$OMEGA

 0.500
 0.312
 0.103
 0.584
```

```
$SIMULATION (123456789) ONLYSIM

$TABLE NOPRINT ONEHEADER FILE=EHR.PAR
  ID TIME DV AMT EVID MDV A1 A2 A3
  ALAG1 D1 CL V1 V2 Q2 K13 T31 IPRED
```

Data Set

#ID	TIME	DV	AMT	CMT	RATE	EVID	MDV
1	0	.	10	1	–2	1	1
1	0.5	.	.	1		0	0
1	1	.	.	1		0	0
1	1.5	.	.	1		0	0
1	2	.	.	1		0	0
1	2.5	.	.	1		0	0
1	3	.	.	1		0	0
1	3.5	.	.	1		0	0
1	4	.	.	1		0	0
1	4.5	.	.	1		0	0
1	5	.	.	1		0	0
1	6	.	.	1		0	0
1	7	.	.	1		0	0
1	8	.	.	1		0	0
1	9	.	.	1		0	0
1	10	.	.	1		0	0
1	14	.	.	1		0	0
1	18	.	.	1		0	0
1	23	.	.	1		0	0
1	24	.	.	1		0	0
1	25	.	.	1		0	0
1	26	.	.	1		0	0
1	28	.	.	1		0	0
2	0	.	10	1	–2	1	1
2	0.5	.	.	1		0	0
2	1	.	.	1		0	0

Pharmacometric Knowledge Discovery from Clinical Trial Data Sets

ENE I. ETTE

14.1 INTRODUCTION

Rational drug development is a model-based, knowledge-driven drug development where the objective is to characterize the response surface—the interplay of drug regimen, exposure, and patient factors to elicit response (efficacy/safety)—that would result in the right dose for the right patient at the time of marketing of the drug. This implies using pharmacometric knowledge discovery and creation (1–4), incorporating population pharmacokinetic/pharmacodynamic (PK/PD) approaches (5–7), clinical trial simulation (8), and appropriate statistical analysis for the characterization of the response surface. Understanding the relationship between drug exposure, response, and patient factors is crucial to rational drug development and pharmacotherapy. This involves extracting the knowledge hidden in clinical data sets to characterize the response surface, thereby defining the utility window for drug therapy.

Pharmacometric knowledge discovery (PMKD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data by characterizing data structure by means of a model (1, 2). PMKD should be implemented in every phase of drug development. This implies that an informative PK/PD sampling design (9–13) is implemented in every study to permit the discovery of knowledge from clinical trial data sets that would give insight into the nature of the dose–concentration–response relationship, and how this is modulated by subject factors. The ultimate use of the implementation of PMKD during drug development is the mapping of the response surface. Defining the process, therefore, will aid the understanding of drug action.

14.2 PHARMACOMETRIC KNOWLEDGE DISCOVERY PROCESS

The purpose of data analysis and interpretation, in general, is to find out, among other things, meaningful patterns and relationships between variables under consideration. It is analysis that transforms data (information) into knowledge. The purpose of a population pharmacometric (PM) data analysis from a population PM data set is to extract knowledge available in the population data set. Sufficient knowledge of the problem under investigation as well as knowledge of the pathophysiology of the disease, pharmacokinetics, mathematics, and statistics are essential elements for a successful PMKD from a population data set. The challenge of performing PMKD on a population data set is to make effective use of the data set to discover the untapped knowledge that lies hidden therein. Specifically, it is the task of implementing (and developing) methods that can discover interesting, useful patterns and relationships that will aid the mapping of the response surface with the results used: (a) to support mission-critical decision making in drug development, (b) to identify and promote the most beneficial drug therapy, (c) for predictions and clinical trial simulations, (d) for the explanation of variability and making dosage recommendations, or (e) to explain drug action in general. These constitute some of the objectives of population modeling.

  Knowledge discovery is an emerging, interdisciplinary research field that lives at the intersection of computer science (database, artificial intelligence, graphics, and visualization), statistics, and several application domains such as clinical pharmacology and pharmacometrics. The details of PMKD have been covered in other works by the author and his collaborators (e.g., see Refs. 1 and 2). Therefore, the intent of this chapter is to provide a brief summary of previous work on the subject and to expand on some key ideas of PMKD. In the sections that follow, the steps taken to perform PMKD are summarized, followed by a discussion of the techniques used in PMKD, challenges in PMKD, an application example, and a summary of the chapter.

14.2.1 Steps in the PMKD Process

The following briefly summarizes the steps taken to perform PMKD in a large clinical trial set:

1. Defining or stating the objective of the PMKD process.
2. Creating a data set on which PMKD will be performed. Data preparation step is a very critical step in the PMKD. Sometimes more effort can be expended in preparing data than in analysis.
3. Data quality analysis (i.e., cleaning and processing the data) (5, 7, 14).
4. Data structure analysis, exploratory examination of raw data (dose, exposure, response, and covariates) for hidden structure, and the reduction of the dimensionality of the covariate vector.
5. Stating assumptions made in the PMKD process.
6. Determining the basic PK model that best describes the data and generating post hoc empiric individual Bayesian parameter estimates.

7. Searching for patterns and relationships between parameters and covariates through graphical displays and visualization.

8. Exploratory modeling using modern statistical modeling techniques such as generalized additive modeling (GAM) (15), cluster analysis, and tree-based modeling (TBM) to reveal structure in the data and initially select explanatory covariates.

9. Consolidating the discovered knowledge into an irreducible form (1), that is, developing a population PM model using the nonlinear mixed effects modeling approach, for example.

10. Determining model robustness through sensitivity analysis, examination of parametric/nonparametric standard errors, and stability testing with or without predictive performance depending on the objective of the PMKD.

11. Interpreting the results: the PMKD process prescribes that the model developed is interpreted in a relational manner. That is, do the findings of the PMKD make sense in the domain in which they will be used? Can the results be communicated in a manner that they can be used? Only if they make sense can the results be considered as "knowledge" (which is viewed pragmatically here).

12. Applying (or utilizing) the discovered knowledge: the pragmatic view of knowledge implies that the results of the PMKD process must have some impact on the way individuals act. Thus, the discovered knowledge must be applied to demonstrate how it can be used.

13. Communicating the discovered knowledge.

These steps are summarized in Figure 14.1. It is important to note that PMKD is an iterative process and the figure therefore is a composite of the steps in the process. For example, modeling encompasses steps 3–10 above.

The PMKD process must be focused. Having a clearly defined objective for the process greatly influences the remainder of the steps in the process. For instance, the choice of data set(s) to be used in the PMKD process is determined by the objective that prompts the process.

In addition, creating a data set on which PMKD is to be performed is not a trivial task. When data are combined across clinical trials, attention must be paid to the variables in the data sets being combined and data completeness (or incompleteness). Data access tools must be available for the pooling of data stored in different data warehouses and platforms.

Generally, not much attention has been given to data quality analysis or structure revelation in a population PM data set, which could provide a link between the data set and the analytical path chosen for PMKD. A lack of these analyses can result in reduced power and the production of biased population PM parameter estimates. In order to avoid these outcomes that can be caused by a nonsystematic or improper data analysis, the above structured approach was first proposed by Ette et al. (1) and reinforced by Williams et al. (2).

Aspects of data quality that need to be analyzed are correctness and completeness. Correctness of concentration–response-time data relative to dosing history and covariate information can be checked by comparing the records in the population PM data set with information in the case report forms, and this can be done by
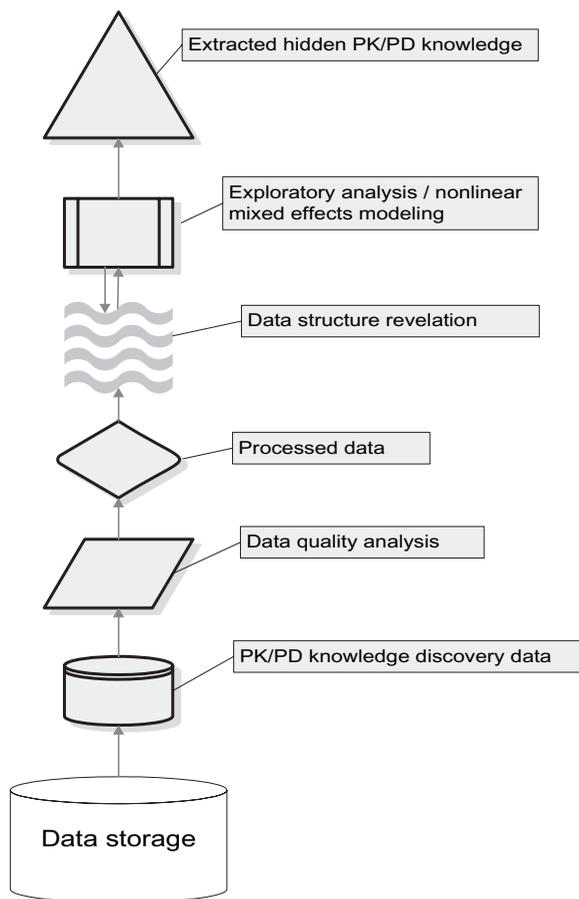
FIGURE 14.1   Overview of a compressed PK/PD knowledge discovery process.

using a sample of the records. Completeness of population PK and PD data records is a property that cannot be satisfied due to many reasons, such as omissions when recording or inputting data into a clinical database, or the malfunctioning of medical equipment (16). Data incompleteness must be addressed before proceeding with the PMKD process. In order to prepare data for population PMKD, some imputation of data may be done, and there are different procedures available for handling missing data (5, 17–19) (see Chapter 9).

Data structure analysis is the examination of the raw data for "hidden" structure, outliers, or leverage observations. This is repeated during the exploratory modeling (and nonlinear mixed effects modeling) steps using case deletion diagnostics (20). This type of analysis is important since outliers or leverage observations may occur in a population PM data set. It is equally important for the reduction of the covariate vector.

The knowledge discovery basis of PM modeling permits the generation of hypotheses from the relationship discovered during data structure analysis. These relationships can be tested in the nonlinear mixed effects modeling step. It can

also suggest a testable hypothesis that can be independently tested via traditional confirmatory experiments and analysis (1).

One of the most difficult tasks for a pharmacometrician/pharmacokineticist is to convey findings from a PMKD process to clinicians and other members of the drug development team. The use of high-quality graphics can effectively enhance the communication of PM knowledge to a drug development or medical research team. Graphics, in particular, are essential for conveying relations and trends in an informal and simplified visual form (1, 20, 21). Failure to communicate these findings successfully puts at risk all the data analysis efforts, irrespective of its quality.

14.3 SOME TECHNIQUES EMPLOYED IN PMKD

In this section some of the techniques for PMKD are discussed. The techniques covered here are only some of the more common techniques used in PMKD, since space would not permit a discussion of all the techniques. Thus, linear and multiple linear regression techniques are not discussed. These can be found in general statistics textbooks. Rather, we focus our attention on visualization, GAM, clustering with emphasis on tree-based modeling (TBM), nonlinear mixed effects modeling, and computer-intensive approaches for characterizing parameter estimation reliability and stability testing.

14.3.1 Visualization

The eye–brain system is the single most powerful information processor available to humans, and effective use is made of this in PMKD. There is no single statistical tool that is as powerful as a graph. Through graphical displays one can put the most effective information processing system to good use to obtain insight into the structure of the data. Graphs can convey an enormous amount of quantitative information. The eye–brain system can not only summarize vast amounts of information quickly and extract salient features, but it is also capable of focusing on detail. There may be many patterns and relationships present even in small data sets that are considerably easier to discern in graphical displays than by any data analytic method.

Visualization, which is graphing and fitting, makes effective use of one's eye–brain system. With visualization one has a penetrating look at the data structure. The knowledge of the subject under study should guide what is learned from the data. When data are visualized effectively there can be a sudden interocular traumatic impact, a conclusion that hits one between the eyes. Thus, visualization is useful for data structure revelation. Visualization has to be combined with statistical inference to help calibrate the uncertainty about an outcome. When this is the case, visualization is useful for checking assumptions.

A pharmacometrician can gain insight from appropriate data displays that is virtually impossible to gain from looking at tables of output or simply summaries of statistics. For some tasks, appropriate visualization is the only process needed to confirm a hypothesis or solve a problem.

There is a wide range of visualization techniques that are appropriate for PMKD. S-Plus (Insightful, Seattle, WA), SAS (SAS Institute, Carey, NC), and Spotfire

(Spotfire, Sommerville, MA) are some of the statistical packages that provide useful visualization graphics to complement their statitiscal/modeling capabilities. These include scatterplot matrices, box plots, conditioning plots, three-dimensional plots, and multidimensional "point cloud" and interactive visualizations such as "brushing." In PMKD the pharmacometrician needs to use visualization, modeling, and statistical analysis iteratively.

14.3.2   Generalized Additive Modeling (GAM)

New techniques for data analysis abound in statistical literature. GAM is a powerful tool technique, and a full historical account of GAM with ample references can be found in the research monograph of Hastie and Tibshirani (15). GAM is closer to a reparameterization of the model than a reexpression of the response. Once an additive model is fitted to the data, one can plot their p coordinate functions separately to examine the roles of predictors in modeling response. With the GAM approach the dependence of a parameter (P) on covariates (predictors) $X_1, \ldots, X_p$ are modeled. Usually, the multiple linear regression (MLR) approach is the method of choice for this type of problem. The MLR model is expressed in the following form:

$$P = \alpha + \sum_{j=1}^{P} \beta_j X_j + \varepsilon \qquad (14.1)$$

where $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. This model makes a strong assumption of the linear dependence of $E(P)$ (the expectation of P or mean response) on the predictors. The MLR model is extremely useful and convenient if this assumption holds, even roughly, because it provides a description of the data, summarizes the contribution of each predictor with a single coefficient, and provides a simple method for predicting new observations.

The assumption of linear dependence of the response variable on each of the predictors may not always hold. For many types of data a change in the mean of the response variable is accompanied by a change in its variance. The GAM approach presents a general perspective for the handling of covariates in a multiple regression setting. The linear form of $\alpha + \Sigma_{j=1}^{P} \beta_j(X_j)$ is replaced with the additive form $\alpha + \Sigma_{j=1}^{P} f_j(X_j)$,

$$P = \alpha + \sum_{j=1}^{P} f_j(X_j) + \varepsilon \qquad (14.2)$$

where $f_j(X_j)$ is an arbitrary univariate function that is either a linear function or a smoothing spline. Since each covariate is represented separately in Eq. (14.2), GAM retains the important interpretive feature of the linear model: the variation of the fitted response surface holding all but one predictor fixed does not depend on the values of the other predictors. In practice, this means that once the additive model is fitted to the data, one can plot the P coordinate functions separately to examine the roles of predictors in modeling the response. The estimated function

forms of GAM are analogs of the coefficients in MLR. Thus, separate functions are introduced to allow for nonlinearity and heterogeneous variances. This is closer to a reparameterization of the model than to a reexpression of the response.

With GAM the "data" (covariate and individual Bayesian PM parameter estimates) would be subjected to a stepwise (single-term addition/deletion) modeling procedure. Each covariate is allowed to enter the model in any of several functional representations. The Akaike information criterion (AIC) is used as the model selection criterion (22). At each step, the model is changed by addition or deletion of a covariate that results in the largest decrease in the AIC. The search is stopped when the AIC reached a minimum value.

Similarly, if the response is a binary variable, (0, 1) or (yes, no), and p is the probability of positive response, then the above equation can be rewritten as follows:

$$\text{logit}(p) = \alpha + \sum_{j=1}^{p} f_j(X_j) + \varepsilon \tag{14.3}$$

Model building for the generalized additive logistic model proceeds in the same manner as described above for GAM with a continuous response variable.

14.3.2.1 Partial Residuals Plot

Residual analysis is of vital importance in any regression analysis. A residual analysis entails the careful evaluation of the differences between the observed values and the predicted values of the dependent variable after fitting a regression model to the data. Residual plots are used interalia with a view to identifying any undetected tendencies in the data, as well as outliers and fluctuation in the variance of the dependent variable (21). However, interpretation of such residual plots requires great care on account of the possible degree of subjectivity involved therein.

Partial residuals are produced with GAM, and not the usual residual plots. Plots of residuals and functions of residuals are useful particularly for identifying patterns in the data that may suggest heterogeneity of variance or bias due to deterministic model misspecification or misspecifications of the regression variables. One particular form of bias that may exist occurs when a predictor variable is included in the model in a linear form when it actually has a curvilinear or nonlinear relationship with the response variable. A plot used by Ezekiel (23) and later referred to as a partial residual plot by Larsen and McCleary (24) is useful for this purpose. Partial residuals are defined as

$$r_i = y_i - \left(\hat{y}_i - \hat{\beta}_j x_{ij}\right) = \varepsilon_i + \hat{\beta}_j x_{ij} \tag{14.4}$$

Since

$$\hat{y}_i = \alpha + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip} \qquad i = 1, 2, \ldots, n \tag{14.5}$$

(where α is the intercept and β is the regression coefficient) it follows that $\hat{y}_i - \hat{\beta}_j x_{ij}$ is an estimate of the ith response when all the predictors except the jth on (x_j) are used, hence the name partial residuals. A plot of r against the predictor x_j thus allows one to examine the relationship between y and x_j after eliminating the influence of other predictors. Figure 14.2, for example, shows partial residual plots of
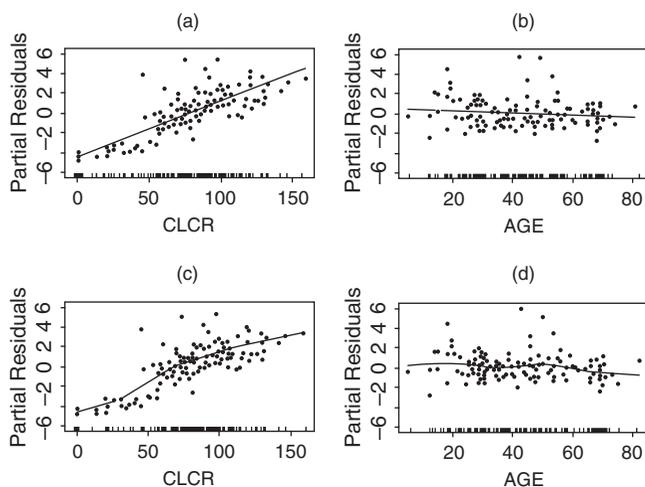
FIGURE 14.2   Scatterplots of partial residuals of clearance [CL (L/h)] of a drug versus (A) creatinine clearance [CLCR (mL/min)] and (B) age (yr) from multiple regression analysis; CL (L/h) versus (C) CLCR (mL/min); and (D) age (yr) from GAM analysis. The same scale is used for the ordinate in each plot so that the relative importance of each covariate can be compared.

the contributions of covariates (creatinine clearance (CLCR) and age) to explaining variability in clearance of a test drug. The inadequacy of the linear model (i.e., multiple linear regression) compared with GAM is obvious.

In addition, in least-squares residual plots of ε_i versus x_j, the slope of the regression line of ε_i against x_j can be expected to be zero. In contrast, the regression of r against x_j should have a slope equal to $\hat{\beta}_j$, the coefficient of x_j when the full model is fitted. This property of partial residuals makes these plots useful in assessing the extent of possible nonlinearity in a certain predictor (25). If the slope of the plot of r against x_j approximately equals the coefficient $\hat{\beta}_j$ obtained from a fit of the full model, the specification of x_j in the regression model can be assumed to be correct.

It is often claimed that partial residual plots are useful omnibus plots that allow detection of outliers, influential or leverage observations, nonlinearity, and other informative nonrandom patterns. The detection of nonlinearity, however, is the central motivation for partial residual plots.

14.3.3   Clustering

Clustering is a descriptive task used to identify a finite set of categories or clusters to describe the data. Also called data segmentation, cluster analysis has many goals that all relate to segmenting and collecting data into subsets or "clusters." Each data point within a cluster is more closely related to each other than those assigned to a different cluster. A response such as blood pressure can be described by a set of measurements, or by its relation to subject variables such as age or renal function. The goal can sometimes be to arrange the data clusters into natural hierarchy. The categories may consist of a richer representation such as overlapping or hierarchical categories, or they may be mutually exclusive and exhaustive. Discovering a

homogeneous subpopulation such as poor metabolizers or subjects with impaired renal function (20) are good examples of clustering applications in a PMKD context. The degree of similarity, or dissimilarity, of data being clustered is central to cluster analysis.

There are many approaches to clustering. A popular approach based on hierarchical groupings partitions (or classifies) data into categories according to a measure of similarity. Starting with as many "groups" as there are observations (one observation per group), the process ends up with one group containing all observations. Depending on their similarity, at each step of the process observations are added to existing groups or become the beginnings of new groups. An example would be a classification "tree" created to describe the classification process graphically. The tree, called a dendrogram, displays the number and types of groups formed at each similarity level. This potentially yields insight into the underlying or hidden data structure. There are many approaches to clustering (see Hastie et al. (26) for an in-depth discussion), but the emphasis in this chapter is on tree-based modeling, which is discussed subsequently.

14.3.4 Tree-Based Modeling (TBM)

TBM is an exploratory modeling technique for uncovering structure in the data and assessing the adequacy of linear models (20, 27). It operates only on ranks of the data. It is this aspect of TBM for a numeric explanatory variable that renders it invariant to monotone transformations of the explanatory variable. It automatically incorporates interactions between covariates such as when one parameter–covariate relationship depends on another covariate. These are important advantages that TBM has over GAM.

14.3.5 Population Modeling

Population PM modeling, herein referred to as population modeling, seeks to identify the measurable pathophysiologic factors that cause changes in the dose–concentration–response relationship and the extent of these changes so that, if such changes are associated with clinically significant shifts in the therapeutic index (i.e., safety margin), dosage can be appropriately modified. Population modeling is discussed in detail in Chapters 4, 8, 10, and 28 in this text. Population modeling seeks to develop an irreducible model (28), given a data set as a summary of extracted knowledge, about an aspect of the response surface contained in the clinical trial data. In doing this, it is important to ensure that the appropriate covariates are retained in the irreducible model, and this can be done via PM model stability testing.

14.3.6 Stability of a PM Model

Without considering the stability of a PM model in an independent sample, it is possible to be unaware of the fact that some factors represent spurious associations with the outcome because of "noise" in the data or multiple comparisons. Furthermore, minor changes in the data set may result in the selection of different covariates. This might leave one in a quandary as to which covariates actually are of predictive importance. When statistical significance is the sole criterion for

including a covariate in the model, the number of variables selected is a function of the sample size.

Since a PM model may be used not only for the explanation of variability but also for predictions (28), being certain about covariates that are retained in the model and the predictive accuracy of the model is important. Thus, the stability of the PM model (in terms of the covariates) and its predictive performance is essential. "Stability" is used in the sense of "replication stability" for inclusion of covariates in a model (29). Sample sizes are usually too small (especially in pediatric studies) to apply the well known and often recommended method of data splitting (30). With better computer facilities, a computer-intensive method such as the related bootstrap method has proved to be a practicable alternative (31) (see Chapter 15 of this text). The method proposed by Ette (31) for stability testing to ensure that appropriate covariates are selected to build a PM model is described below.

One hundred bootstrap samples are generated and the appropriate structural model that best describes the data from each sample is determined. This is done to ensure that the model that best describes the bootstrap data is not different from the basic structural model used for developing the population PK model for the data before bootstrapping. With the right structural model POSTHOC individual Bayesian estimates are generated and the "data" subjected to GAM.

For each bootstrap replication a selection method is used to identify the significant variables. Important predictive covariates should be included in the selected model in most bootstrap replications, since it is assumed that each replication, being a random sample from the patients in the study, should reflect the underlying structure of the data. Therefore, the percentage of inclusion in the model is a criterion for the predictive importance of a covariate. Where there is only one candidate covariate to be included in the model, there is a direct relationship between the selection level and the bootstrap inclusion fraction. Applying the bootstrap technique to the Cox regression model, Sauerbrei and Schumacher (32) showed that for a significance level of $\alpha = 0.05$, the bootstrap inclusion fraction is 0.50 for each bootstrap replication. It was proved that the inclusion of a covariate at a selection level of $\alpha = 0.05$ in the original data can be based equivalently on a cutoff value for the bootstrap inclusion fraction using a selection level of $\alpha = 0.05$ in each replication. This criterion has been proposed for the selection of a covariate among the covariates obtained from GAM for the final NONMEM replication stability step discussed subsequently.

A really important covariate should be entered into the model in nearly all of the bootstrap replications (31). Thus, the strategy is to select covariates that will be useful in explaining variability in the PM parameter of choice.

The approach developed by Ette (31) for the determination of model stability is summarized in the following steps:

Step 1. Determine the basic PM model for the characterization of population pharmacometrics using nonlinear mixed effects modeling.

Step 2. Generate 100 bootstrap samples, each having the same sample size as the original data set, using nonparametric bootstrap.

Step 3. Apply GAM to each of the 100 bootstrap replicates with a selection level of $\alpha = 0.05$ and a frequency cutoff value of 0.50. Those covariates that do not attain the cutoff value should be eliminated.

Step 4. With the appropriate pharmacostatistical models, perform nonlinear mixed effects modeling to develop a PM model using covariates retained in step 3 with a covariate selection level of $p \leq 0.005$. Backward elimination for covariate selection should be applied to each of the 100 bootstrap samples. The covariates found to be important in explaining variability in the parameter of interest should be used to build the final (irreducible) population model. The choice of $p < 0.005$ is to indirectly take the multicomparisons that would be made into account.

Step 5. The PM model developed in step 4 should then be applied to the original data set to obtain PK parameters for the drug. Confidence intervals can be constructed for the parameters of the model using asymptotic standard errors of estimates. A confidence interval for a regression quotient incorporating zero for any particular covariate is suggestive of further studies to investigate the importance of that covariate in the characterization of the pharmacometrics of the drug, although the covariate could not be part of the population PM model developed for the data set analyzed.

The publicly available S-Plus macro, Xpose (33), can be used to automatically implement the GAM aspect of stability testing.

14.3.7 Reliability of Estimates

The reliability of the parameter estimates can be checked using a nonparametric technique—the jackknife technique (20, 34). The nonlinearity of the statistical model and ill-conditioning of a given problem can produce numerical difficulties and force the estimation algorithm into a false minimum.

The preciseness of the primary parameters can be estimated from the final fit of the multiexponential function to the data, but they are of doubtful validity if the model is severely nonlinear (35). The preciseness of the secondary parameters (in this case variability) are likely to be even less reliable. Consequently, the results of statistical tests carried out with preciseness estimated from the final fit could easily be misleading—thus the need to assess the reliability of model estimates. A possible way of reducing bias in parameter estimates and of calculating realistic variances for them is to subject the data to the jackknife technique (36, 37). The technique requires little by way of assumption or analysis. A naive Student t approximation for the standardized jackknife estimator (34) or the bootstrap (31, 38, 39) (see Chapter 15 of this text) can be used.

14.4 SOME CHALLENGES IN PMKD

14.4.1 High Dimensionality

With enhanced computational power it is now possible to work with very large data sets. This means dealing with a large number of records in the database and a large number of variables, with the consequence being that the dimensionality of the problem is high. The problem with high-dimensional data sets is that there is an increase in the size of the search space for model induction. In addition, it increases the probability of finding spurious patterns that are not valid. Reducing the effective

dimensionality of the problem through structure revelation to eliminate collinear variables and the use of prior knowledge to identify irrelevant variables are ways to overcome this problem.

14.4.2 Missing Data

Important attributes may be missing if the database was not designed with PMKD in mind. A possible solution is to use multiple imputations or other imputation techniques as dictated by the type of "missingness" to create a complete data set for PMKD (17–19, 40, 41).

14.4.3 Time-Varying Covariates

Time-varying covariates can provide additional information to that obtained from time-constant covariates in characterizing variability, if properly accounted for in population modeling. Higgins et al. (42) proposed a two-step approach, while Wahlby et al. (43) proposed two alternate approaches—(a) splitting the standard covariate model into a baseline covariate effect and a difference from baseline covariate effect, and (b) allowing the magnitude of the covariate effect to vary between individuals by including interindividual variability in the covariate effect— for handling time-varying covariates. Readers are referred to the appropriate references for details.

14.5 APPLICATION EXAMPLE

14.5.1 PMKD Objective and Data

The objective was to develop a descriptive population PK model for a test drug. A drug was administered to 88 full-term infants and children in six studies (31). There were 48 boys and 40 girls. Thirteen patients were human immunodeficiency virus (HIV) seropositive. The patients' mean weight was 22.15 ± 18.27 kg, mean age was 6.78 ± 4.30 years, and the average serum creatinine value was 0.80 ± 0.31 mg/dL. The reciprocal of serum creatinine (RSC) was used for population PK modeling. The patients received either a single dose or multiple doses of drug orally or intravenously. An average of 11.1 (range, 1–21) concentrations were measured per participant.

14.5.2 PMKD for Population Pharmacokinetic Model Development

The structured approach for PMKD from a population PK data set outlined in Section 14.2.1 was used for population PK model development. A two-compartment linear PK model with a step function was used to describe the data. The step function was used because of a change in clearance of drug (with a resultant change in concentration (Figure 14.3)) after 1 week as a consequence of the maturation process occurring in the elimination organs. The focus of this example is on a stable population PK model that was developed for the estimation of clearance (CL) for the test drug. Table 14.1 summarizes the results from steps 2 and
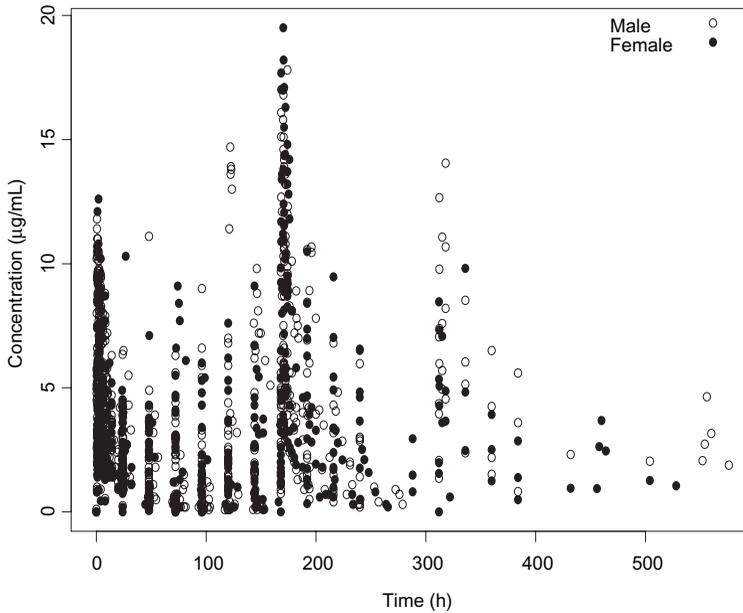
FIGURE 14.3 Concentration–time plot for males (open circles) and females (closed circles).

TABLE 14.1 Selection of Covariates

Regression Method	Covariate	Selection Frequency
Generalized additive modeling	Sex	0.30
	RSC	0.73
	HIV status	0.70
NONMEM	RSC	0.45
	HIV status	0.23

NONMEM, nonlinear mixed effects modeling; RSC, reciprocal of serum creatinine; HIV, human immunodeficiency virus.

3 above. GAM yielded sex, RSC, and HIV status as predictors of CL. However, the final NONMEM population PK model for the drug incorporated only RSC as a linear predictor of CL. The final population PK model for CL and its variability was described with the following equations:

```
If (TIME. LE. 180) then
TVCL = THETA(1)*WT**THETA(2)+THETA(5)RSC
Else
TVCL = THETA(3)*WT**THETA(4)+THETA(6)RSC
CLj = TVCL*EXP(ηjCL),
```

where TVCL was the typical population value for CL, CL_j is CL in the jth patient, η_{jCL} represents the intersubject variability, and the thetas are intercepts and regression

coefficients. HIV status was found to be a significant predictor ($p < 0.05$) in the initial model building step but was not statistically significant in the final model development step in which the level of significance was set at $p < 0.005$ to compensate for multiple comparison. Table 14.2 summarizes the results of the model developed without the replication stability step.

Given the frequency cutoff value of 0.50 at $\alpha = 0.05$, RSC and HIV status were selected by GAM (Table 14.1). Although gender was selected by GAM in some bootstrap samples, it did not meet the criteria for the retention of a covariate in the subsequent NONMEM population model building step. Because a more conservative significance level ($p < 0.005$) was chosen for the NONMEM replication stability step, the criteria applied to the GAM step were not applied here. Consequently, RSC and HIV were retained for the final NONMEM model (Table 14.2). These covariates were then used in fitting the model to the original data set (without replication). The results are summarized in Table 14.3. The imprecision associated with the estimation of the regression coefficient on HIV status is worth noting. The 95% confidence interval (95% CI) on this coefficient was (−0.215, 0.729). Also, other parameter estimates obtained from the population model stability step were similar to those obtained without replication stability. The incorporation of zero in the 95% CI for the regression coefficient on HIV status confirmed the fact that the final population PK model obtained before the replication stability evaluation was stable. It was noted that although HIV status was not included in the final model,

TABLE 14.2   Summary of Final Population Pharmacokinetic Model Parameter Estimates with Reference to Clearance: No Replication

Parameter	Population Value (SE)
THETA(1)—Intercept for CL (L/h) for time ≤180 h	0.293 (0.065)
THETA(2)—WT on CL for time ≤180 h	0.837 (0.174)
THETA(3)—Intercept for CL (L/h) for time >180 h	0.269 (0.112)
THETA(4)—WT on CL >180 h	2.000 (0.030)
THETA(5)—RSC on CL for time ≤180 h	0.010 (0.003)
THETA(6)—RSC on CL >180 h	0.153 (0.051)

CL, clearance; WT, weight; RSC, reciprocal of serum creatinine.

TABLE 14.3   Replication Stability: Summary of Final Population Pharmacokinetic Model Parameter Estimates with Reference to Clearance

Parameter	Population Value (SE)
THETA(1)—Intercept for CL (L/h) for time ≤180 h	0.297 (0.055)
THETA(2)—WT on CL for time ≤180 h	0.832 (0.137)
THETA(3)—Intercept for CL (L/h) for time >180 h	0.262 (0.086)
THETA(4)—WT on CL >180 h	2.070 (0.223)
THETA(5)—RSC on CL for time ≤180 h	0.010 (0.001)
THETA(6)—RSC on CL >180 h	0.141 (0.045)
THETA(7)—HIV on CL for time ≤180 h	0.257 (0.236)

CL, clearance; WT, weight; RSC, reciprocal of serum creatinine.

the selection of a covariate from a stability step of model building (although not in the final model) is a strong indication for further investigation of the covariate in explaining variability in the pharmacokinetics of the drug studied. This outcome of the PMKD process was appropriately communicated to the team.

14.6 SUMMARY

A structured approach for PMKD from a clinical trial data set is described. The process is an iterative one, involving heterogeneous tasks. The approach in this chapter lays out systematically how hidden knowledge can be discovered from a clinical trial data set with the use of modern graphical, modeling, and statistical approaches. These techniques for PMKD have been applied to a PK data set as a demonstration of how PMKD can be performed. The techniques described in this chapter give the pharmacometrician the liberty to choose a pharmacostatistical methodology appropriate to the problem at hand with the maximization of knowledge extraction, rather than on the basis of mathematical/statistical tractability.

REFERENCES

1. E. I. Ette, P. J. Williams, H. Sun, E. O. Fadiran, F. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).

2. P. J. Williams, C. J. Godfrey, A. Roy, H. M. Chu, and E. I. Ette, Pharmacokinetics/pharmacodynamics knowledge discovery and creation during drug development, in *Industrial Pharmacokinetics*, P. Bonate and D. Howard (Eds.). AAPS Press, Arlington, VA, 2004 pp. 163–202.

3. E. I. Ette, H. M. Chu, and C. J. Godfrey, Data supplementation: a pharmacokinetic/pharmacodynamic knowledge creation approach for characterizing an unexplored region of the response surface. *Pharm Res* **22**(4):523–531 (2005).

4. E. I. Ette and H. M. Chu, Pharmacokinetic/pharmacodynamic knowledge creation: toward characterizing an unexplored region of the response surface, in *Pharmacometrics: The Science of Quantitative Pharmacology*, E. I. Ette and P. J. Williams (Eds.). Wiley, Hoboken, NJ, 2007, Chapter 32.

5. *Guidance on Population Pharmacokinetics for Industry.* Food and Drug Administration, Rockville, MD, 1999.

6. H. Sun, E. O. Fadiran, C. D. Jones, L. Lesko, S. M. Huang, K. Higgins K, C. Hu, S. Machado, S. Maldonado, R. Williams, M. Hossain, and E. I. Ette, Population pharmacokinetics: a regulatory perspective. *Clin Pharmacokinet* **37**:41–58 (1999).

7. P. J. Williams and E. I. Ette, The role of population pharmacokinetics in drug development in the light of the FDA Guidance on population pharmacokinetics. *Clin Pharmacokinet* **39**:385–395 (2000).

8. H. C. Kimko and S. Duffil S (Eds.), *Clinical Trial Simulations.* Marcel Dekker, New York, 2002.

9. E. I. Ette, H. Sun, and T. M. Ludden, Design of population pharmacokinetic studies, *Proc Am Stat Assoc (Biopharmaceutics Sec)*: 487–492 (1994).

10. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional pharmacokinetic studies: implications for pediatric and animal studies. *Clin Res Regul Affairs* **13**(3&4):133–165 (1996).

11. E. O. Fadiran, C. D. Jones, and E. I. Ette, Designing population pharmacokinetic studies: performance of mixed designs. *Eur J Drug Metab Pharmacokinet* **25**:231–239 (2000).

12. A. Roy and E. I. Ette, A pragmatic approach to the design of population pharmacokinetic studies. *AAPS J* **7**(2):E408–420 (2005).

13. E. I. Ette and A. Roy, Designing population pharmacokinetic studies for efficient parameter estimation, in *Pharmacometrics: The Science of Quantitative Pharmacology*, E. I. Ette and P. J. Williams (Eds.). Wiley, Hoboken, NJ, 2007, Chapter 12.

14. T. H. Grasela, E. J. Antal, J. Fiedler-Kelley, D. J. Foit, B. Barth, D. W. Knuth, B. J. Carel, and S. R. Cox, An automated drug concentration screening and quality assurance program for clinical trials. *Drug Info J* **33**:273–279 (1999).

15. T. Hastie and R. Tibshirani, *Generalized Additive Models*. Chapman & Hall, London, 1990.

16. E. Shortliffe and L. E. Perreault (Eds.), *Medical Informatics: Computer Applications in Health Care*. Addison-Wesley, Reading, MA, 1990.

17. D. B. Rubin, Inference and missing data. *Biometrika* **63**:581–582 (1976).

18. D. B. Rubin, Multiple imputation after 18+ years. *J Am Stat Assoc* **91**:473–489 (1996).

19. E. I. Ette, H. M. Chu, and A. Ahmad, Data imputation, in *Pharmacometrics: The Science of Quantitative Pharmacology*, E. I. Ette and P. J. Williams (Eds.). Wiley, Hoboken, NJ. 2007, Chapter 9.

20. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

21. E. I. Ette, Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial. *Ann Pharmacother* **32**:818–828 (1998).

22. H. Akaike, Information theory and an extension of the maximum likelihood principle, in *Second International Symposium on Information Theory*, B. Petrov and F. Csaki (Eds.). Akademia Kiado, Budapest, 1973, pp. 267–281.

23. M. A. Ezekiel, Method of handling curvilinear correlation for any number of variables. *J Am Stat Assoc* **19**:431–453 (1924).

24. W. A. Larsen and S. J. McCleary, The use of partial residual plots in regression analysis. *Technometrics* **14**:781–789 (1972).

25. R. F. Gunst and R. L. Mason, *Regression Analysis and Its Applications*. Marcel Dekker, New York, 1980.

26. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York, 2001.

27. L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, 1984.

28. E. I. Ette, P. J. Williams, Y. H. Kim, L. R. Lane, M.-J. Liu, and E. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).

29. J. Gifi, *Nonlinear Multivariate Analysis*. Wiley, Chichester, UK, 1990.

30. D. R. Cox, A note on data-splitting for the evaluation of significance levels. *Biometrika* **62**:441–444 (1975).

31. E. I. Ette, Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* **37**:486–495 (1997).

32. W. Sauerbrei and M. Schumacher, A bootstrap resampling procedure for model building: application to Cox regression model. *Stat Med* **11**:2093–2109 (1992).

33. E. N. Jonsson and M. O. Karlson, Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Meth Prog Biomed* **58**:51–64 (1998).

34. D. V. Hinkley, Jackknife confidence limits using Student *t* approximations. *Biometrika* **64**:21–28 (1977).

35. N. R. Drapper and H. Smith, *Applied Regression Analysis*. Wiley, Hoboken, NJ, 1966, pp. 263–304.

36. M. H. Quenouille, Notes on bias estimation. *Biometrika* **43**:353–360 (1956).

37. R. G. Miller, The jackknife—a review. *Biometrika* **61**:1–15 (1974).

38. B. Efron, Nonparametric standard errors and confidence intervals (with discussion). *Can J Stat* **9**:139–172 (1981).

39. E. I. Ette and L. C. Onyiah, Estimating inestimable standard errors in population pharmacokinetic studies: the bootstrap with winsorization. *Eur J Drug Metab Pharmacokinet* **27**:213–224 (2002).

40. P. R. Rosenbaum and D. B. Rubin, The central role of the propensity score in observational studies for causal effect. *Biometrika* **70**:41–55 (1983).

41. P. W. Lavori, R. Dawson, and D. Shera, A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med* **14**:1913–1925 (1995).

42. K. M. Higgins, M. Davidian, and D. M. Giltinan, A two-step approach to measurement error in time-varying covariates in nonlinear mixed-effects models, with application to IGF-1 pharmacokinetics. *J Am Stat Assoc* **92**:436–448 (1997).

43. U. Wahlby, A. H. Thomson, P. A. Milligan, and M. O. Karlsson, Models for time-varying covariates in population pharmacokinetic–pharmacodynamic analysis. *Br J Clin Pharmacol* **58**:367–377 (2004).

Resampling Techniques and Their Application to Pharmacometrics

PAUL J. WILLIAMS and YONG HO KIM

15.1 INTRODUCTION

Pharmacometric (PM) models have many and varied applications for drug development, regulation, and applied pharmacotherapy. Resampling techniques can be applied to model development, evaluation, and validation—most often resulting in an economy of effort once applied to these aspects of modeling (1–3). Models have been defined as either descriptive or predictive (see Chapter 8). While descriptive models require checks for reliability and stability, predictive models have the added requirement of validation (which resampling can do).

Traditionally, resampling has been used for PM model (PMM) covariate selection, bias correction (the difference between the estimator of a parameter, $\hat{\theta}$, and the parameter, θ), the estimation of standard errors, the construction of confidence intervals, and model validation (1–3, 5–8). Most often validation has been thought of as the ability of a model to make external predictions (4). Obtaining an external data set can be time consuming, costly, and difficult. For some populations, such as pediatric patients or patients with rare conditions, obtaining a validation data set may be nearly impossible. Resampling methods present the opportunity to validate a model internally. Thus, the index data may be used to validate the model, saving the effort or difficulty of obtaining an external validation data set. Cross-validation, bootstrapping, and the posterior predictive check have all been used as internal validation procedures (1, 4–11).

Given that resampling when applied to PMMs can result in greater confidence in estimated models, economy of effort, and improved models, a review of the techniques and their application would be profitable.

Jackknife (JKK), cross-validation, and the bootstrap are the methods referred to as resampling techniques. Though not strictly classified as a resampling technique, the posterior predictive check is also covered in this chapter, as it has several characteristics that are similar to resampling methods.

15.2 RESAMPLING AND THE PLUG-IN PRINCIPLE

In general, resampling is based on the plug-in principle. The plug-in principle states that an estimate of a parameter, $\theta = t(F)$, is defined to be $\hat{\theta} = t(\hat{F})$, where θ is the true parameter from F, the true probability distribution, and $\hat{\theta}$ is the empirical estimate of θ from $\hat{F}$, the empirical distribution of F: that is, one estimates a property of F such as summary statistics by using $\hat{F}$. In general, the plug-in principle works well unless there is information about F not provided by the sample ($\mathbf{x}$).

15.3 DESCRIPTIVE SUMMARIES OF RESAMPLING METHODS

15.3.1 The Jackknife

M. H. Quenouille introduced the jackknife (JKK) in 1949 (12) and it was later popularized by Tukey in 1958, who first used the term (13). Quenouille's motivation was to construct an estimator of bias that would have broad applicability. The JKK has been applied to bias correction, the estimation of variance, and standard error of variables (4, 12–16). Thus, for pharmacometrics it has the potential for improving models and has been applied in the assessment of PMM reliability (17). The JKK may not be employed as a method for model validation.

The JKK can be used for any estimator that is a sample analog of a parameter. For instance, one can use the JKK for the sample mean as an estimator of the population mean, the sample variance as an estimator of the population variance, the sample minimum as an estimator of the population minimum, and so on. This definition can be extended to any population characteristic and is therefore of interest in pharmacometrics, especially when applied to population modeling.

The JKK is a direct application of the plug-in principle. To understand the JKK, let us denote the estimator of θ by $\hat{\theta}$, where $\hat{\theta}$ is based on a sample of size n. The JKK estimator $\hat{\theta}_{\text{JKK}}$, of v is defined as follows. Calculate n estimators $\hat{\theta}_{(i)}$, where for each $i = 1$ to n, $\hat{\theta}_{(i)}$ is obtained using the expression defining $\hat{\theta}$ *eliminating the i*th *observation* so that each $\hat{\theta}_{(i)}$ is calculated with a sample of size $n - 1$. Each observation is removed once from the data and the procedure of interest is carried out on the data. For this reason, the JKK is often also known as the *leave-one-out* method. If one now defines the mean of the $\hat{\theta}_{(i)}$, $i = 1, \ldots, n,$ as

$$\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{(i)} \tag{15.1}$$

The JKK estimate of bias is

$$\text{Bias}_{\text{JKK}}\left(\hat{\theta}\right) = (n-1)\left(\hat{\theta}_{(\cdot)} - \hat{\theta}\right) \tag{15.2}$$

The use of the JKK estimate of bias for bias correction is presented later in this chapter.

In 1958, J. Tukey (13) proposed a JKK estimate for the variance of any sample analog estimator $\hat{\theta}$. This can be written as

$$Var_{JKK}(\hat{\theta}) = \frac{(n-1)}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \tag{15.3}$$

and hence the JKK estimate of the standard error of $\hat{\theta}$ is simply

$$\widehat{SE}_{JKK}(\hat{\theta}) = \sqrt{Var_{JKK}(\hat{\theta})} \tag{15.4}$$

An important aspect of any inference is the construction of approximate confidence intervals for $\hat{\theta}$. A parameter is distributed as a Student's t distribution with $(n-1)$ degrees of freedom. An approximate $100(1-\alpha)\%$ confidence interval for θ is given by

$$\left[\hat{\theta}_{JKK} - t_{1-\alpha/2}(n-1) \bullet \widehat{SE}_{JKK}(\hat{\theta}), \hat{\theta}_{JKK} + t_{1-\alpha/2}(n-1) \bullet \widehat{SE}_{JKK}(\hat{\theta}) \right] \tag{15.5}$$

where $t_{1-\alpha/2}(n-1)$ denotes the $(1-\alpha/2)$ quantile of a Student's t distribution on $(n-1)$ degrees of freedom. Unfortunately, this approach to confidence interval construction has been documented to work poorly.

   The alternative to the leave-one-out approach to the JKK is the grouped or blocked JKK (18). Here there are g blocks of size s. The grouped JKK can save time by executing the PM procedure on the g blocks. Here again, one has an estimator of θ, $\hat{\theta}_b$, which is the estimate of θ with the block eliminated. Next, a pseudo value $(P_{n,s,-j})$ is calculated as follows:

$$P_{n,s,-j} = g\hat{\theta} - (g-1)\hat{\theta}_b \tag{15.6}$$

Then the JKK estimate of θ (θ^*_{JKK}) is

$$\theta^*_{JKK} = \frac{1}{g} \sum_{i=1}^{g} P_{n,s,-j} \tag{15.7}$$

and the estimated variance (V_{JKK}) of θ^*_{JKK} is

$$V_{JKK} = \frac{1}{g(g-1)} \sum_{i=1}^{g} (P_{n,s,-j} - \theta^*_{JKK})^2 \tag{15.8}$$

   For all approaches to the JKK the advantage of working with $\hat{\theta}_{JKK}$ is that there is no need for Monte Carlo simulation of the index data. The disadvantage is that there is usually increased error of estimation (15). Efron (16) has demonstrated that estimates of standard errors are better when the bootstrap is applied compared to the JKK. Furthermore, the JKK can fail if the statistic or parameter $\hat{\theta}$ is not smooth. By smooth one means that small changes in the data result in small changes in $\hat{\theta}$. The most typical nonsmooth statistic is the median. Often in PM data analyses and model development statistics or parameters are nonsmooth because data sets are small, containing leverage observations, and the algorithms for converging onto parameters result in nonsmooth $\hat{\theta}$ values. Therefore, one must exercise caution when applying the JKK for the purpose of checking PMM reliability or for bias correction.

15.3.2 Cross-Validation

The JKK is a process that focuses on statistical accuracy, bias correction, standard errors, and confidence intervals estimation. In contrast, cross-validation (CV) is an old idea that follows a similar algorithm to the JKK but focuses on predictability via prediction error generation. Prediction error aims at assessing the ability of the model to predict some future observation, usually a dependent variable. Although this method is often considered to be the same as repeated data splitting, it has some advantages over data splitting because when using CV the size of the model development database can be much larger so that less data are discarded from the estimation process and one does not rely on a single sample split, which increases variability (15, 16).

Of interest here is *how well a model will predict some future response* in some external or new data. One may look at the average residual error in the data set from which the prediction rule was estimated, called the apparent error rate. However, this estimate of the residual error (apparent error here) will be too optimistic and will underestimate the true prediction error. The problem here is that the training and assessment sample are the same. CV is used to correct this underestimation of the apparent error rate.

There are two types of CV: *leave-one-out* and grouped (*K*-fold) methods. In both processes the data are divided into a training set and an assessment set. In the *leave-one-out* approach, one subject (or one data point) is omitted from the training set, and the remaining $N - 1$ subjects or data (this is the training set) are used to estimate the model. Since this chapter is focused on resampling for PM modeling, the emphasis is on subjects and not data points. Then the prediction rule of the model is fixed and one predicts into the omitted observation (the assessment data) and the prediction error is estimated. The modeling process is repeated N times with each subject being omitted from the estimated model. The true error rate of this model is estimated to be the average of all the error rates from the N test models.

In the *grouped* method, one divides the data into K approximately equal-sized groups (e.g., into 10 to 20 equal-sized groups), and the data-splitting method is repeated K times. These data should be well mixed to avoid distorting the results of the CV when the grouped approach is taken. Each time, one of the K subsets is used as the assessment set and the other $K - 1$ subsets are put together to form a training set. The prediction rule of the model is fixed to that estimated from the $K - 1$ subset and one predicts into the assessment group of observations and the average prediction error is estimated. Next, using all the data, the final model is created. The true error rate of this model is estimated to be the average of all the error rates from the K test models (5, 19). A schematic representation of K-fold CV is presented in Figure 15.1.

Simulation experiments have demonstrated that CV is on average unbiased. However, simulation has also shown that when the data from a CV exercise are rearranged, the results are highly variable (19–21). To minimize or overcome this problem, the data should be well mixed.

15.3.3 Bootstrapping

Bradley Efron (16) first suggested bootstrapping in 1979 as a tool for constructing inferential procedures in modern statistical data analysis. Resampling has been
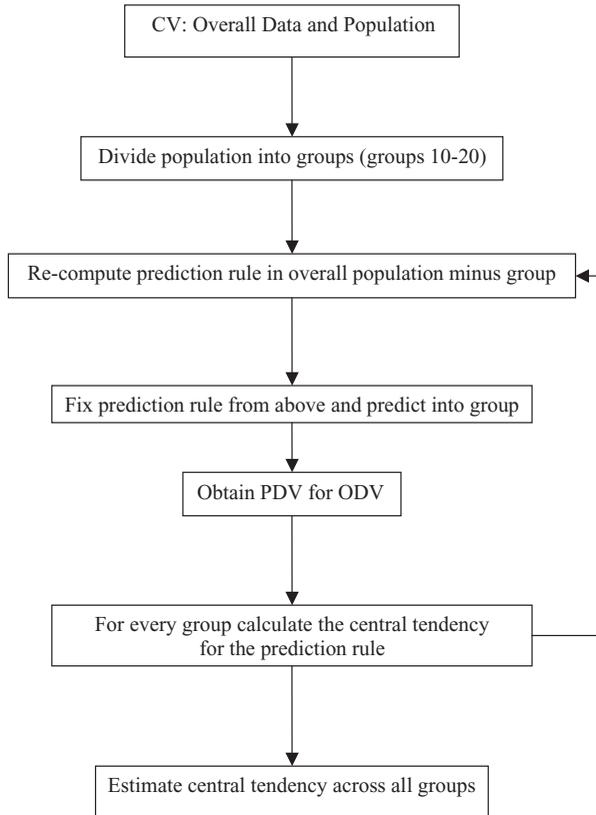
FIGURE 15.1 Internal validation: cross-validation flowchart (PDV is the predicted dependent variable, ODV is the observed dependent variable).

defined as a method of repeatedly generating pseudosamples distributed according to the same distribution as the original sample. The procedure of interest is then carried out on each pseudosample and then the results of the application of these procedures to the pseudosamples are summarized (14).

This approach allows theorem proving with minimal data collection and carrying out its functions without underlying assumptions about the distribution of the data. The bootstrap has had wider application than JKK or CV to pharmacometric modeling. It has been used for model building, for estimating statistical moments that could not be estimated by applying parametric approaches, for comparing non-hierarchical models, and for model evaluation and validation (1–3, 7, 8, 17, 22, 25). However, it is computationally intense—a drawback that has become unimportant with the development of computers of high speed that have large virtual memory. Several different versions of the bootstrap have been proposed (23, 24, 26).

15.3.3.1   *Standard (Nonparametric) Bootstrap*
In this version, no assumptions, such as a normal distribution, are made about the index (original) data. It is executed generally as follows (14, 24):

1. The original data are assumed to be an independent and identically distributed sample of size m, from an unknown probability distribution, $G \rightarrow (x_1, x_2, \ldots,$

x_n). The empirical distribution function, $\hat{G}$, of this sample is the discrete distribution that sets the probability of each value x_i ($i = 1, 2, \ldots, m$) in the sample to the value of $1/m$.

2. A bootstrap sample is generated by repeated random sampling, with replacement, of an m-sized "pseudosample" from the original data set. At each sampling step, every vector x_i has an equal probability of being chosen. Thus, for a given iteration, it is possible to choose three of x_1, none of x_2, five of x_3, and so forth.

3. This sampling is repeated until the bootstrap sample also consists of m vectors, $Y^* = (x_1^*, x_2^*, \ldots, x_i^*, \ldots, x_m^*) \sim \hat{G}$, where the vector x_i^* represents all the observations for the ith randomly selected subject.

4. Each set of pseudodata is used and the procedure of interest is applied to each pseudodata set.

5. Tabulation and summarization of the results of the application of the procedure to each of the pseudosamples is done.

The nonparametric maximum likelihood (NPML) method is a nonparametric bootstrap because $\hat{F}$ is the nonparametric estimate of F (14). The NPML concept states that "given a set of unknown terms and a set of data related to the unknowns, the best estimate of the unknowns consists of the values that render the set of data most probable." A schematic representation of the bootstrap is presented in Figure 15.2.

Bootstrapping is the resampling with replacement method that has the advantage of using the entire data set. It has been demonstrated to be useful in PMM validation (1, 3, 22) and has the same advantages as do other internal validation methods in that it obviates the need for collecting data from a test population. Bootstrapping has been applied to population pharmacokinetic (PPK) model development, stability check and evaluation, and bias estimation (1–3, 25).
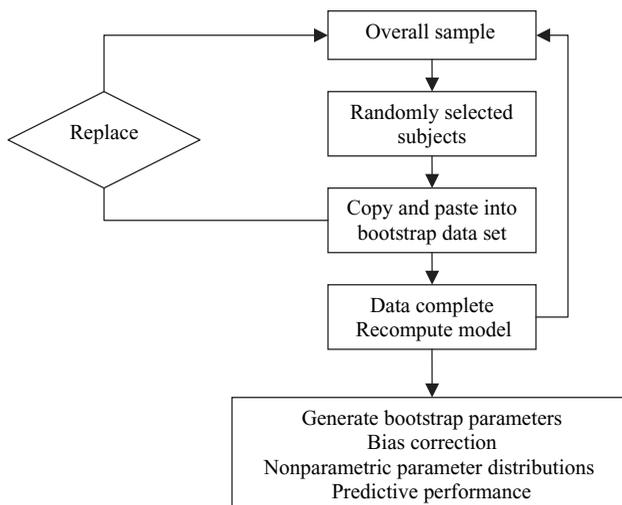
FIGURE 15.2 Bootstrap flow.

15.3.3.2 Bootstrapping Residuals

The most common approach to constructing bootstrap pseudosamples is to bootstrap the pairs: that is, one randomly selects data for a typical data set on a line-by-line or subject-by-subject basis that is inserted into the pseudosample and replaced. Bootstrapping residuals is another approach that has particular application to regression analyses. In a typical bootstrap data set (B_i), data are chosen of the form

$$B_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \ldots, (x_{in}, y_{in})\}$$

When the residuals are bootstrapped, the B_i are of the form

$$B_i = \{(x_{i1}, x\hat{\beta} + \hat{\varepsilon}_{i1}), (x_{i2}, x\hat{\beta} + \hat{\varepsilon}_{i2}), \ldots, (x_{in}, x\hat{\beta} + \hat{\varepsilon}_{in})\}$$

where $\hat{\varepsilon}_{in}$ is the ith residual. It can be seen here that the residuals from the regression are bootstrapped and added to the predicted values of the dependent variables.

Bootstrapping the residual assumes that the residuals are not a function of the dependent variables and that the form of the error model is known. This is a strong assumption that is seldom met in regression analyses and pharmacometrics in particular. Bootstrapping pairs is less sensitive to assumptions than is bootstrapping residuals.

15.3.3.3 Smoothed Bootstrap

The smoothed bootstrap has been proposed to deal with the discreteness of the empirical distribution function $(\hat{F})$ when there are small sample sizes $(N < 15)$. For this approach one must smooth the empirical distribution function and then bootstrap samples are drawn from the smoothed empirical distribution function, for example, from a kernel density estimate. However, it is evident that the proper selection of the smoothing parameter (h) is important so that oversmoothing or undersmoothing does not occur. It is difficult to know the most appropriate value for h and once the value for h is assigned it influences the variability and thus makes characterizing the variability terms of the model impossible. There are few studies where the smoothed bootstrap has been applied (21, 27, 28). In one such study the improvement in the correlation coefficient when compared to the standard nonparametric bootstrap was modest (21). Therefore, the value and behavior of the smoothed bootstrap are not clear.

15.3.3.4 Parametric Bootstrap

For the smoothed bootstrap the shape of the distribution is not assumed. However, if one assumes F to be continuous and smooth, then the next step is to assume that it has a parametric form. If one assumes that F has a parametric form such as the Gaussian distribution, then the appropriate estimator for F would be a Gaussian distribution.

For the parametric bootstrap instead of resampling with replacement from the data, one constructs B samples of size n, drawing from the parametric estimate of $\hat{F}_{par}$. Here $\hat{F}_{par}$ is the parametric distribution of F. The procedures of interest are then applied to the B samples in the same manner as for the nonparametric bootstrap.

However, in parametric problems the bootstrap adds little or nothing to the theory or application and one cannot explain why the typical approach to estimating parameters via formulas should be replaced by bootstrap estimates. Consequently, it is uncommon to see the parametric bootstrap used in real problems. When applied to population pharmacometric (PPM) modeling, a weakness of the parametric bootstrap is that it assumes that the model is known with a high degree of certainty. This is seldom true.

15.3.3.5 Double Bootstrap
The double bootstrap was a method originally suggested by Efron (15) as a way to improve on the bootstrap bias correction of the apparent error rate of a linear discrimination rule. It is simply a bootstrap iteration (i.e., taking resamples from each bootstrap resample). The double bootstrap has been useful in improving the accuracy of confidence intervals but it substantially increases computation time and most likely increases the incidence of unsuccessfully terminated runs. It has been applied to linear models but not to PM modeling.

15.3.3.6 Bayesian Bootstrap
The Bayesian bootstrap was introduced by Rubin (26) in 1981 and subsequently used by Rubin and Schenker (29) for multiple imputation in missing-data problems. The Bayesian bootstrap is not covered because its application is for multiple imputation of missing data and this is addressed in Chapter 9.

15.3.3.7 Bootstrap Standard Error Estimates
The bootstrap is a very useful procedure when one wishes to estimate the standard error (SE) of a parameter (θ) from an unknown probability distribution (F). The original introduction of the bootstrap was for the purpose of estimating the SE of $\hat{\theta}$.

To execute the nonparametric bootstrap SE ($S\hat{E}_B$) the following process must be executed:

1. Select B independent bootstrap samples. This will usually be at least 100 bootstrap data sets for a PM model (1).
2. Perform the evaluation of interest on each bootstrap sample, estimating the parameter of interest from each sample.
3. Estimate the SE of the parameter of interest by the sample standard deviation from B bootstrap samples:

$$\widehat{SE}_B = \left\{ \sum_{i=1}^{n} \left[\hat{\theta}(b) - \hat{\theta}*(\cdot) \right]^2 \Big/ (B-1) \right\}^{1/2} \tag{15.9}$$

where $\hat{SE}_B$ is the bootstrap standard error, $\hat{\theta}(b)$ is the estimate of θ from the bth pseudosample, and $\hat{\theta}*(\cdot)$ is the mean $\hat{\theta}$ across all B pseudosamples.

4. Once the SE has been calculated, one should always display the histogram of the results of the bootstrap replications.

15.3.3.8 Bootstrap Confidence Intervals

Standard errors are most often used to assign confidence intervals to parameters of interest. Given an estimated parameter $(\hat{\theta})$ and an estimated standard error $(\widehat{SE})$ the usual 95% confidence interval is

$$\hat{\theta} \pm 1.960\widehat{SE} \tag{15.10}$$

where 1.960 comes from the standard normal table. The above is valid only as $n \to \infty$; therefore, the studentized t distribution provides a better approximation:

$$\hat{\theta} \pm t_{(n-1)}\widehat{SE} \tag{15.11}$$

where $t_{(n-1)}$ is Student's t distribution on $n - 1$ degrees of freedom.

The use of the t interval does not account for skewness or kurtosis. An alternative method, called the bootstrap percentile confidence intervals, is less dependent on assumptions and therefore less affected by these factors. Furthermore, very often one makes transforms of a θ to normalize the distribution, but an appropriate transform is not always apparent. The percentile method can be thought of as an algorithm for automatically incorporating these transforms. The only assumption that one makes with the percentile method is that an appropriate transform exists, which does not need to be known.

15.3.3.9 Bootstrap Bias Estimation

$\hat{\theta}$ is an estimate of θ from the data at hand, $\hat{F}$. $\hat{\theta}$ will always be a biased estimate of θ and we are interested in obtaining the least biased estimate of θ. A large bias is an undesirable characteristic of a parameter.

Bias is the difference between an estimator $\hat{\theta}$ and the true quantity θ. In addition to the JKK, the estimation of bias can also be done by application of the bootstrap. Bias estimation using the bootstrap is discussed here while the use of bias estimates for bias correction is addressed later.

The bootstrap estimate of bias is

$$\widehat{\mathrm{Bias}_b} = \hat{\theta}(\cdot)_b - \hat{\theta} \tag{15.12}$$

where $\hat{\theta}(\cdot)_b$ is the mean of the parameter from all the bootstrap replications and $\hat{\theta}$ is the estimate of the parameter from the original data. It is important to note that when the ratio of the estimated bias is small relative to the standard error of the parameter $(\widehat{SE}_B)$, then the bias is of little concern.

Jones et al. (25) have investigated bias correction via Monte Carlo simulation(s). The study design was cross sectional with a profile (block) randomized sampling time design, which appeared to result in an inflation of the estimate of the interindividual random effect for apparent volume of distribution. The authors attempted to correct this positively biased parameter by applying the bootstrap. The authors bootstrapped both pairs and weighted residuals (WRES). Bootstrapping the WRES is attractive because it assumes that the error in the data set does not depend on the concentration: that is, the residuals are interchangeable. Of note, the resampling of pairs on a subject-by-subject basis resulted in less bias correction for the interindividual random effect associated with apparent volume than the resampling

of the weighted residuals. When the weighted residuals were resampled, the bias correction for this parameter was 15%. Of further importance was the fact that regardless of the approach to resampling, estimates were stable at 5000 or more bootstrap estimates.

15.3.3.10 Bootstrap Estimates of Prediction Error

It is often of interest to estimate the predictability of the dependent variable when the prediction rule is applied to a similar set of external data (data from which the model was not estimated). However, obtaining a similar set of data is often costly or impossible; therefore, some internal estimate of external predictability would be advantageous.

It is intuitive that the predictability of the dependent variables into the training data set from which a model was estimated will be optimistic, when compared to predicting into an external data set. In such a case, the prediction errors will have a downward bias. Therefore, a method that estimates predictability for external data is needed and this can be executed via the bootstrap.

The most commonly used parameter to assess predictability for the bootstrap has been the squared prediction error (SPE). The SPE refers to the square of the difference between a future response and its prediction from the model:

$$PE = E(ODV - PDV) \tag{15.13}$$

and therefore

$$SPE = E(ODV - PDV)^2 \tag{15.14}$$

where ODV and PDV are the observed and predicted dependent variables, respectively, and E (the expectation) refers to repeated sampling from the true population.

Bootstrapping provides estimates of predictive precision that are nearly unbiased and that are of relatively low variance. When each of the models is applied to its own bootstrap data set, there is for each ODV a PDV that is generated. The most common prediction error operator is the mean squared prediction error (MSPE). For our notation here, $MSPE(M_1, D_1)$ is an estimate of the MSPE when model 1 is applied to bootstrap data set 1. For example, if there were 200 bootstrap data sets, this would be done until $MSPE(M_1, D_1)$ to $MSPE(M_{200}, D_{200})$ were obtained using 200 bootstrap data sets (1, 2, 4, 14, 20). Ette (1) showed that 200 bootstrap replicates was adequate for the determination of the predictive performance of a model.

The next step in assessing predictive accuracy is to apply the frozen models (M_1 to M_{200}) to the data set D_0, which is the original data set (not from bootstrap resampling) with all individuals occurring once. In this process, the coefficients and random effects are fixed for M_1 to M_{200}, so that at this step one has $MSPE(M_1:D_0)$ to $MSPE(M_{200}:D_0)$.

The next step is to estimate the optimism (OPT) (or bias due to overfitting in the final model fit) for the prediction operator. This is executed for each model so that

$$OPT_i = MSPE(M_i:D_0) - MSPE(M_i, D_i) \tag{15.15}$$

where OPT_i is the optimism for the ith model. It is expected that $MSPE(M_i, D_i)$ will be smaller than $MSPE(M_i:D_0)$ because $MSPE(M_i, D_i)$ is making predictions into the data set from which it was estimated and $MSPE(M_i:D_0)$ is making predictions into the original data set, which serves as an "external" data set because M_i is a PMM from a bootstrap sample.

Next, the mean OPT across all the models is estimated and in this case is

$$OPT = \frac{1}{n}\sum_{i=1}^{n} OPT_i$$

(15.16)

Once the mean optimism is known, it is added to the results of the prediction operator that was estimated from the original data set. This results in an improved estimate of the prediction operator as the prediction operator estimates generated when the process is applied to its own data will be overly optimistic compared to applying the prediction operator to the universe of possible external data sets.

$$\overline{MSPE}_{imp} = MSPE(M_0:D_0) + OPT$$

(15.17)

where $MSPE_{imp}$ is the improved estimate of the $MSPE$ provided by the bootstrap and M_0D_0 is the $MSPE$ when the original model was applied to the original data. The $MSPE_{imp}$ is estimated to provide an estimate of the $MSPE$ that would result if the model were applied to an external data set. The lower the $MSPE_{imp}$, the better the model.

15.3.3.11 *Model Building with the Bootstrap*
Ette (1) provides an example of the application of bootstrapping to PMM building, specifically to a population pharmacokinetic (PPK) model. In this study it was desired that the deterministic model (one-compartment versus two-compartment) and the covariates for inclusion be known with a high degree of certainty (1, 3).

Conceptually, for each bootstrap replication a selection method is used to identify the significant covariates and the deterministic model. Important predictive covariates should be included in most bootstrap replications, as it is assumed that each replication should reflect the underlying data structure. Therefore, an important covariate should be included in nearly all of the bootstrap replications.

Application of bootstrapping to this process proceeded as follows:

1. One hundred nonparametric bootstrap data sets were generated.
2. Using NONMEM® (University of California at San Francisco), each data set was fit to a one- and two-compartment model. A two-compartment model fit the data better for all bootstrap data sets.
3. NONMEM was used to estimate the parameters for each bootstrap data set. Individual Bayesian parameters were generated. These estimates along with covariates formed a new data set.
4. Generalized additive modeling (GAM) was applied to each of the output data sets. A selection criteria of $\alpha = 0.05$ and a frequency cutoff of 0.50 was applied for continued investigation of a covariate; that is, GAM had to select a covariate for inclusion in 50% or more of the models from the 100 bootstrap fits for the covariate to be considered for further investigation.

5. Those covariates that did not attain the cutoff level were eliminated from further investigation.
6. A full model was constructed and applied to each of the bootstrap data sets with backward selection and $\alpha = 0.005$ for retention in the final model. Those variables found to be important in the model were retained to build the final PPK model.
7. The population model with the proper deterministic model and covariates was then applied to the original data to obtain parameter estimates.

Using this approach, several covariates were excluded from further study at step 4 above and two covariates were retained. The final model was determined expeditiously and most importantly with a high degree of certainty.

15.3.3.12   Comparing Nonhierarchical Models Using the Bootstrap

Until recently, no method of comparing nonhierarchical regression models has been available. The bootstrap has been proposed because it may estimate the distribution of a statistic under weaker conditions than do the traditional approaches. In general, for nonlinear mixed effects models that are not hierarchical, the preferred model has simply been selected as that with the lower objective function (2). A more rational approach has been proposed for comparing nonhierarchical models, which is an extension of Efron's method (2, 30). The test statistic is the difference between the objective functions (log-likelihood difference—LLD) of the two non-hierarchical models. The method consists of constructing the confidence interval for the LLDs.

To execute this, an estimate of the sample distribution of the LLD under the null hypothesis must be derived to perform a test. The bootstrap method for estimating sample distribution of the difference of the objective function given the observations is used to solve the problem. This allows one to reject the null hypothesis of equal noncentrality parameters, that is, of equality of fit if zero is not contained in the confidence interval so derived. One thousand bootstrap pseudosamples were constructed, the nonhierarchical models of interest were applied, and the percentile method for computing the bootstrap confidence intervals was used.

15.3.3.13   Estimating Inestimable Standard Errors

PM parameters are most often estimated by assuming asymptotic normality. Often the standard errors or confidence intervals for PM parameters are not accurately estimated because distributions are heavily tailed or skewed or contain influence data. In addition, when sample sizes are small, it is impossible to obtain accurate or precise standard errors of parameters.

Another problem that can occur is when the 95% confidence intervals cross 0 and values below 0 make no sense. For example, one may obtain estimates of a random effect, say, the coefficient of variation for a parameter, and along with this the standard error for the coefficient of variation. Sometimes when asymptotic normality is assumed, the 95% confidence interval for the coefficient of variation can be less than 0 at its lower bound. It does not make sense to have a coefficient of variation for a parameter that is less than 0. The bootstrap can be used to avoid this error and estimate confidence intervals that make sense. An example of applying bootstrapping to deal with the concerns described here can be found in a paper by Ette and Onyiah (8).

15.3.4 Posterior Predictive Check (PPC)

D. B. Rubin first suggested PPC in 1984 (31) as a tool for constructing inferential procedures in modern statistical data analysis. In this approach a model is estimated directly from the index data, and then a new set of data is generated through the simulation of the resulting model. The simulated data set is compared with the index data to see if the model's deficiencies have a noticeable effect on the substantive inferences (9). The basic approach for PPC within the context of PPK modeling is as follows:

Step 1. Derive an irreducible (i.e., with natural uncertainty) model from a probable set of data.

Step 2. Estimate the model parameters from the final irreducible model, given the data.

Step 3. Simulate the dependent variable from the final irreducible model to obtain several replicates of the data.

Step 4. Estimate the posterior model and obtain fixed and random effects from the model. This is a plausible model from the posterior distribution of the dependent variables obtained in step 3.

Step 5. From the posterior model simulate data where the dependent variable $(DV) = y^{\text{rep}}$. This is a plausible set of dependent variables from a plausible model in step 4.

Step 6. Perform statistics on these dependent variables; for example, this involves reestimating model parameters to get root mean-squared error (RMSE) and mean absolute error (MAE).

Step 7. Compare statistics from the replicated data sets to the original data set. A schematic representation of posterior predictive check is shown in Figure 15.3.
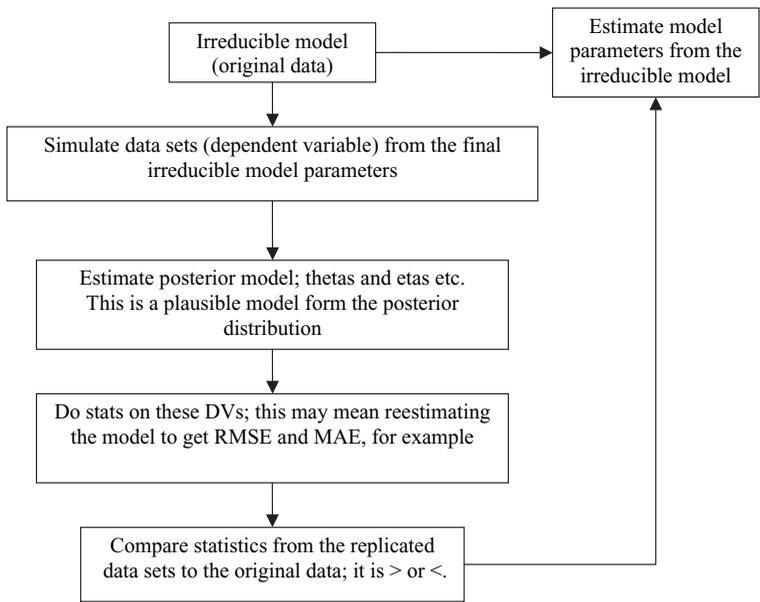
FIGURE 15.3   Schematic representation of the posterior predictive check flow.

Few papers have been published that apply the PPC to PM modeling (9–11, 32). In a limited evaluation of PPC, the authors concluded that the failure of the PPC to invalidate a model offers little assurance that the model is correct (32). Further evaluation of the utility of the PPC, especially in hierarchical models, is needed.

15.4   BIAS CORRECTION

The estimation of bias and bias correction can be done by application of the bootstrap and the JKK. Bias correction is discussed here and the reader is referred to Sections 15.3.1 and 15.3.3.9. A large bias is an undesirable characteristic of a parameter.

The bias corrected parameter ($\bar{\theta}$) is calculated from

$$\bar{\theta} = \hat{\theta} - \widehat{\text{Bias}} \tag{15.18}$$

There is a wrong tendency to think of $\hat{\theta}(\cdot)_b$ or $\hat{\theta}(\cdot)_{\text{JKK}}$ as the bias-corrected parameter. Note, for example, that if $\hat{\theta}(\cdot)_b$ is greater than $\hat{\theta}$, then the bias-corrected estimate ($\hat{\theta}$) will be less than $\hat{\theta}$.

Bias correction can be dangerous in practice because of the high variability in its estimate. In spite of this, its estimation is usually worthwhile. If bias is small relative to the standard error of the parameter, then it is best to use $\hat{\theta}$ rather than $\bar{\theta}$. If the bias is large compared to the SE of the parameter, then it is an indication that $\hat{\theta}$ is not an appropriate estimate of the parameter θ.

15.5   EXAMPLE OF MODEL EVALUATION AND VALIDATION

The following was taken from a drug development case where data was collected in Phase 1 and Phase 2 and a PPK model was developed to aid in proposing a dosing strategy for a Phase 3 confirming study. There were 323 subjects with 2352 measured concentrations that were used to develop the PPK model. The strategy set forth as model appropriateness (see Chapter 8) was employed as the approach to model development.

15.5.1   Execution of the Example

From the original data set (D_0), 505 bootstrap data sets were constructed by resampling with replacement. The sampling was repeated until each bootstrap sample consisted of N subjects, where in this case $N = 323$. (See Table 15.1 for an extensive explanation of the notation that follows.) These 505 bootstrap data sets were denoted as D_1 to D_{505}. The structure (S_0) of the model M_0 ($M_0 = F(S_0)$) was retained. What was meant by retaining the structure for M_0 was that the coefficients that related the PK parameters (i.e., clearance and apparent volume) to covariates (i.e., weight) were allowed to be estimated from each of the 505 bootstrap data sets. So that for this study, the following was the structural model (S_0):

$$\text{TVCl} = \theta_1 * \text{Wt}$$

TABLE 15.1   Mathematical Notation Utilized in Bootstrapping

Notation	Representation of Notation
D_0	Study data or original data; with each subject's data appearing once and only once.
S_0	The structural model or the developed population pharmacokinetic model with the study data can be expressed as follows: $$TVC1 = \theta_1 * Wt$$ $$TVVd = \theta_2 * Wt$$ $$TVKa = \theta_3$$
F	The fit operator: $M = F(S, D)$ means M is the model produced when structure S is fit to data set D.
D_i	Bootstrapped data or samples (had 323 subjects ($i = 1, 2, 3, \ldots, 323$) subjects' data) (D_0), on which the developed population pharmacokinetic model was based) were drawn with replacement from the observed data (D_0); observed data could either appear in the bootstrap samples (D_i) once, more than one time, or not at all. For each bootstrap data set, the structure was retained but the coefficients and the intercept were reestimated.
$M_i = F(S_0, D_i)$	Models generated from fitting structural model (S_0) to the ($i = 1, 2, 3, \ldots, 505$) bootstrapped samples; bootstrapped models.

Note: 505 bootstrap samples were generated for convenience. For standard nonparametric bootstrap, 200 replicates is adequate.

$$TVVd = \theta_2 * Wt$$

$$TVKa = \theta_3$$

with the between-subject random effects (estimated for clearance (η_1) and apparent volume (η_2) only) and residual random effects modeled as proportional to the parameter of interest. TVCl was the typical value for clearance, TVVd was the typical value for apparent volume, and TVKa was the typical value for the first-order absorption rate constant.

When S_0 was fit (F) to D_i, the model that resulted would be noted as M_i so that when S_0 was applied to D_{11} the model that resulted was noted as M_{11} ($M_{11} = F(S_0, D_{11})$). Thus, there were 505 bootstrap models (M_1 to M_{505}) that were fit, one for each bootstrap data set. For each of the bootstrap data sets it is necessary to fit the structural model to demonstrate it reflects the underlying structure of the data, which is a basic assumption of the bootstrap. Any bootstrap data set that does not support the structural model should be discarded.

For each model fit to the 505 data sets, there were a set of coefficients (θs), that related the covariate to the parameter of interest and random effects. These were compiled in a table and the mean parameters were estimated and compared to those from the original fit.

From the original model (M_0) ($\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, $\hat{\eta}_1$, $\hat{\eta}_2$, and $\hat{\varepsilon}$) and the bootstrap mean parameter estimates ($\hat{\theta}_1(\cdot)_b$, $\hat{\theta}_2(\cdot)_b$, $\hat{\theta}_3(\cdot)_b$, $\hat{\eta}_1(\cdot)_b$, $\hat{\eta}_1(\cdot)_b$, and $\hat{\varepsilon}(\cdot)_b$), the bias was calculated in the general format of $\overline{Bias} = \hat{\theta}(\cdot)_b - \hat{\theta}$. (Note the bias was calculated for η parameters as well.) The 95% confidence intervals were estimated from the

standard nonparametric bootstrap as there were insufficient replicates to employ the percentile method.

The structural model, S_0, was fit to each bootstrap data set. That is, S_0 was fit to D_1 to D_{505}, resulting in models 1 to 505 (M_1 to M_{505}). When each of these models for the bootstrap data sets (D_1 to D_{505}) were estimated, the squared prediction error (SPE) for each concentration was estimated.

$$SPE_{ijk} = (\text{Measured concentration}_{ijk} - \text{Predicted concentration}_{ijk})^2 \qquad (15.19)$$

where SPE_{ijk} was the ith squared prediction error in the jth patient for the kth run, Measured concentration$_{ijk}$ was the ith measured concentration in the jth patient for the kth run, and Predicted concentration$_{ijk}$ was the ith predicted concentration in the jth patient for the kth run. The mean SPE for each bootstrap run was estimated. That is, the model was applied to the data from which it was derived and the individual SPEs were estimated and the mean of the individual SPEs was calculated for each run. These mean errors that were estimated when the M_1 to M_{505} were applied to D_1 to D_{505} were called $Aeboot_1$ to $Aeboot_{505}$ (apparent error, which is the error estimate when a model is applied to the data from which it was estimated).

Next, the models, including the coefficients, were fixed (M_1 to M_{505}) and applied to the original data set (D_0), where each subject appeared once and only once. That is, the model was then applied to a data set from which it was not derived for predictions into that data set. Again for each measured concentration in D_0 there was a predicted concentration and the squared error for each predicted measured concentration pair was estimated using Eq. (15.19). The mean of these SPEs was calculated for each run. These mean errors estimated when M_1 to M_{505} were applied to D_0 were called $PEOrig_1$ to $PEOrig_{505}$.

In a further step, the $Aeboot_1$ to $Aeboot_{505}$ was subtracted from each $PEOrig_1$ to $PEOrig_{505}$, respectively, to estimate a variable called optimism (OPT).

$$OPT_i = PEOrig_i - Aeboot_i \qquad (15.20)$$

This optimism represented the underestimation of the squared prediction error that was expected to occur when the model was applied to the data from which it was derived. In a final step, the average optimism across all bootstrap iterations was estimated and added to the SPE estimated when the M_0 was applied to D_0. This resulted in an improved estimate of the absolute prediction error (SPE_{imp}).

15.5.2   Results

Convergence was successful for 502 of the 505 bootstrap data sets. Three data sets persisted in terminating with rounding errors despite the application of several sets of starting parameters. The results of the bootstrap parameter estimates are presented in Table 15.2 and compared to the results from the PPK model building process. There is strong evidence that the model is without substantive deficiencies and should be accepted as the final irreducible model.

Mean SPE for $M_0:D_0$ was 19.52 and optimism for mean SPE was 2.22. Therefore, the improved SPE (SPE_{imp}) was $SPE_{imp} = 19.51 + 2.22 = 21.73$. In this case, the optimism of the SPE is small compared to the original SPE, again

TABLE 15.2   Bootstrap Estimates for Various Population Pharmacokinetic Parameters Compared to the NONMEM Generated Parameters

Method	θ_1	θ_2	θ_3	η_1	η_2	ε
NONMEM	0.000646	0.0583	0.352	0.112	0.127	0.023
Bootstrap	0.000642	0.0582	0.348	0.110	0.147	0.024
95% CI bootstrap	0.000639, 0.000645	0.0579, 0.0585	0.345, 0.352	0.106, 0.114	0.125, 0.169	0.020, 0.027
Bias/standard error	0.190	0.160	0.169	0.137	0.268	0.171

indicating that the model is without substantive error. Thus, the model was validated.

15.5.3   Conclusions

A PPK model was developed to be used to construct a dosing strategy for a Phase 3 study and therefore needed some form of validation. To obtain a test data set would have been expensive and time consuming. The bootstrap was used to validate the model by estimating the bias/SE and the optimism for the dependent variable. The optimism was small when compared to the SPE of the original sample, thus validating the model.

15.6   SUMMARY

The role of resampling methods in PMM development and validation are explored. If applied, these techniques may bring efficiency to pharmacometric model development and result in models for which one's confidence level is very high. Patient pharmacotherapy will also be improved. One can expect to see resampling more extensively applied to modeling in the future.

REFERENCES

1. E. I. Ette, Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* **37**:486–495 (1997).
2. E. I. Ette, Comparing non-hierarchical models: application to non-linear mixed effects modeling. *Comput Biol Med* **26**:505–512 (1996).
3. E. I. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M. J. Liu, and E. V. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).
4. *Guidance for Industry: Population Pharmacokinetics.* US Department of Health and Human Services, Food and Drug Administration, Rockville MD, February 1999.
5. T. W. Schnider, R. Gaeta, W. Brose, C. F. Minto, K. M. Gregg, and S. L. Shafer, Derivation and cross-validation of pharmacokinetic parameters for computer controlled infusion of lidocaine in pain therapy. *Anesthesiology* **84**:1043–1050 (1996).

6. C. A. Hunt, G. H. Givens, and S. Guzy, Bootstrapping for pharmacokinetic models: visualization of predictive and parameter uncertainty. *Pharmaceut Res* **15**:690–697 (1998).

7. A. Yafune and M. Ishiguro, Bootstrap approach for constructing confidence intervals for population pharmacokinetic parameters: II. A bootstrap modification of standard two stage method for phase I trial. *Stat Med* **18**:601–612 (1999).

8. E. I. Ette and L. C. Onyiah, Estimating inestimable standard errors in population pharmacokinetic studies: the bootstrap with Winsorization. *Eur J Drug Metab Pharmacokinet* **27**:213–224 (2002).

9. A. Gelman, X. L. Meng, and H. Stern, Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**:733–807 (1996).

10. S. B. Duffull, S. Chabaud, P. Nony, L. Chritian, P. Girard, and L. Aarons, A pharmacokinetic simulation model for ivabradine in healthy volunteers. *Eur J Phamaceut Sci* **10**:285–294 (2000).

11. S. B. Duffull and L. Aarons, Development of a sequential linked pharmacokinetic and pharmacodynamic simulation model for ivabradine in healthy volunteers. *Eur J Pharmaceut Sci* **10**:275–284 (2000).

12. M. H. Quenouille, Approximate tests of correlation in time series. *J R Stat Soc B* **11**:18–84 (1949).

13. J. W. Tukey, Bias and confidence in not quite large samples. *Ann Math Stat* **29**:614 (1958).

14. B. Efron and G. Gong, A leisurely look at the bootstrap, the jackknife and cross-validation. *Am Statistician* **37**:36–48 (1983).

15. B. Efron, Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* **78**:461–470 (1983).

16. B. Efron, Bootstrap methods: another look at the jackknife. *Ann Stat* **7**:1–26 (1979).

17. E. I. Ette and T. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**(12):1845–1855 (1995).

18. D. V. Hinkley, Jackknife confidence limits using Student *t* approximation. *Biometrica* **64**:21–28 (1977).

19. M. Stone, Cross-validation choice and assessment of statistical predictions. *J R Stat Soc* **36**:111–147 (1974).

20. M. Stone, An asymptotic equivalence of choice of model by cross validation and Akaike's criterion. *J R Stat Soc* **39**:44–47 (1977).

21. B. Efron, The jackknife, the bootstrap and other reasmpling plans. *SIAM*, Philadelphia abstract #199 (1982).

22. G. W. Boswell, D. R. Miles, P. A. Thiemann, and M. Mesfin, Popuulation pharmacokinetics and bioavailability of motexafin gadolinium (Xcytrin®) in CD1 mice following intravenous and intraperitoneal injection. *Invest New Drugs* **24**:281–289 (2006).

23. B. Efron and R. J. Tibshriani, *An Introduction to the Bootstrap.* Chapman & Hall, New York, 1993.

24. M. R. Chernick, *Bootstrap Methods a Practitioners Guide.* Wiley, Hoboken, NJ, 1999.

25. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional population pharmacokinetic studies: implications for pediatric and animal studies. *Clin Res Reg Affairs* **13**:133–165 (1996).

26. D. B. Rubin, The Bayesian bootstrap. *Ann Stat* **9**:130–134 (1981).

27. E. J. Bedrick and J. R. Hill, A generalized bootstrap, in *Exploring the Limits of Boostrap*, R. Lepage and L. Billard (Eds.). Wiley, Hoboken, NJ, 1992, pp. 319–326.

28. L. Sun and D. Muller-Schwarze, Statistical resampling methods in biology: a case study of beaver dispersal patterns. *Am J Math Manage Sci* **16**:463–502 (1996).

29. D. B. Rubin and N. Schenker, Imputation, in *Encyclopoedia of Statistical Sciences, Update Volume 2*, S. Kotz, C. B. Read, and D. L. Banks (eds.), Wiley, Hoboken, NJ, 1998, pp. 336–342.

30. B. Efron, Comparing non-nested linear models. *J Am Statist Assoc* **79**:791–803 (1984).

31. D. B. Rubin, Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat* **12**:1151–1172 (1984).

32. H. Yano, S. L. Beal, and L. B. Sheiner, Evaluating pharmacokinetic-pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodynam* **28**: 171–191 (2001).

Population Modeling Approach in Bioequivalence Assessment

CHUANPU HU and MARK E. SALE

16.1 INTRODUCTION

Population modeling, called nonlinear mixed effect modeling in statistical terminology, is much more complex in structure than linear or nonlinear regression. The reason is that allowing random effects into nonlinear models opens up a variety of possible models. Typically, it is not a priori clear which model is exactly the best to use, and neither would the data allow sufficient power to distinguish among all possible models. Particularly, in usual applications, the model structure, especially that of the random effects, is not predetermined. For this reason, population modeling is mostly used for the purpose of exploratory data analysis, which is hypothesis generating. Common uses include describing the current data or attempting to best predict unobserved outcomes or relationships.

However, circumstances arise where one may desire to use modeling for confirmatory analyses, as will be discussed later. Analyses of this type are hypothesis confirming, which are inferential in nature. If multiple tests are conducted, adjustment usually must be made to prevent inflation of against Type I error. Therefore, for modeling to be used in confirmatory analyses, special care must be taken to protect against Type I error. Our purpose is to draw attention to this issue, through discussion of two separate application areas in bioequivalence.

16.2 BIOEQUIVALENCE OVERVIEW

16.2.1 Pharmacokinetic Bioequivalence

The most common purpose of bioequivalence (BE) assessment is to evaluate the comparability of bioavailability of drug formulations. Standard BE studies are 2×2 crossover, and subjects are densely sampled so that individual

pharmacokinetic (PK) exposure parameters, AUC and $C_{\max}$, can be determined with precision. These individual PK parameters are then log-transformed to be used with standard models, adjusting for sequence and period effects, to obtain a 90% confidence interval of log(test/reference ratio). The confidence interval is then back-transformed to the original scale to obtain the confidence interval for the test/reference ratio of average PK parameters. Confidence intervals falling within (0.80, 1.25) indicate bioequivalence. For the statistical theory, see Schuirmann (1) for details.

There are also other BE-type assessments. Interaction studies assess the influence on bioavailability by other individual factors, such as food, alcohol, or other drugs. Such studies are usually single-sequence crossover, but the assessment method remains the same—whether confidence intervals of AUC and $C_{\max}$ ratios fall within (0.80, 1.25). The same can be said of PK similarity assessments between subject populations, for example, healthy volunteers versus patients. The assessment method is the same as that used for BE, but important differences remain. In typical BE studies, subjects are densely sampled so that individual PK parameters, AUC and $C_{\max}$, can be determined with precision. PK similarity assessments are concerned with the differences in different populations, instead of formulations. The assessments are usually based on multiple (parallel) studies, as crossover studies are not possible, and sequence and period effects are not considered. Assessments involve obtaining estimates of average PK parameters in the populations and the 90% confidence intervals for the ratios of the average PK parameters.

16.2.2 Pharmacodynamic Bioequivalence

For locally acting drugs, such as pulmonary and topical drugs, the assessment based on PK exposures AUC and $C_{\max}$ is not appropriate. In such situations, plasma concentrations may be irrelevant to efficacy and even unavailable. In the case of metered dose inhalers (MDIs), the US Food and Drug Administration has been basing BE assessment on the dose-scale approach (2), which assesses the relative bioavailability (F) of the test and reference drug based on a pharmacodynamic (PD) endpoint, that is, the fraction of the test product dose that causes the same response in the pharmacodynamic endpoint as one dose of the reference product does. Approval of abbreviated new drug applications (ANDAs) would be based on the 90% confidence interval of F.

Both PK and PD BE assessments concerned here are of the "average BE" type, which are based on population averages. Population modeling is potentially applicable for such assessments, by producing estimates of population average parameters and their confidence intervals. Although standard software generated standard errors could be indicative, they are only approximate for nonlinear models. Therefore, more accurate confidence interval construction methods need to be considered. More difficulty lies in the fact that model selections are typically involved with population modeling and thus contain certain exploratory aspects. On the other hand, BE assessments involve specific hypothesis tests and thus are confirmatory tasks.

We first discuss BE assessment in the presence of sparsely sampled subjects and then discuss pharmacodynamic endpoint bioequivalence.

16.3   METHODS FOR ASSESSING PK BIOEQUIVALENCE WITH PRESENCE OF SPARSELY SAMPLED SUBJECTS

At various stages of drug development, PK similarity assessment may be needed in situations where some or all subjects are sparsely sampled so that evaluating individual PK parameters is not possible. This may happen in some patient studies where sampling is limited for practical reasons or in pediatric studies due to limitation of blood draws. The situation also arises during "bridging" analyses that aim to establish PK similarity between different populations so that safety data in one group could be indicative of another. In such situations, PK similarity assessments must be made from information at hand. When individual PK parameters cannot be determined directly, using model-based approaches seems a meaningful alternative. The population modeling approach seems particularly suitable because it gives a framework allowing the ratios of average AUC and $C_{\max}$ to be predicted from model parameters.

Population modeling has been reported as an alternative tool to examine evidence of BE (3–6). Typically in such investigations, one examines data from some 2×2 crossover BE studies, constructs a population PK model using traditional model building practices (and in some instances, removing outliers), computes the confidence intervals generated from standard error of parameter estimates, and finds the results similar to those given by standard BE assessments. However, there are complexities when using population modeling, a traditional tool for exploratory analysis, for confirmatory analyses such as BE.

16.3.1   Primary Complexity

The subject population covariate plays a special role in PK similarity assessment, similar to that of formulation in BE assessment. That is, if the influence of formulation on the rest of the parameters (K_a, V, CL, etc.) is not adequately represented in the model, then the model will underrepresent the formulation influence on the predictions of PK parameters and may bias the BE assessment results. To further illustrate this, the conventional model building procedure might find the formulation factor "insignificant" in the model, and if the final model contains no formulation factor, it will predict the AUC and $C_{\max}$ ratios to be 1 with certainty, that is, producing confidence intervals of length 0. Thus, BE would have to be declared by default. This is clearly unacceptable from the standpoint of traditional BE assessment, which often finds the formulation term insignificant in ANOVA but always produces confidence intervals of positive lengths for the AUC and $C_{\max}$ ratios.

16.3.2   Additional Complexities

1. Conventional PK model building usually involves much model exploration (e.g., the influence of many potential covariates), whereas traditional BE assessment usually does not pretest any covariates. However, traditional BE assessment generally avoids preliminary tests, for the purpose of protecting against Type I error, that is, mistakenly concluding BE (7). From a statistical

perspective, complex model exploration suffers many deficiencies, including overly optimistic variances of parameter estimates and goodness of fits (8).

2. The model built for BE assessment could depend on the modeler, which may cast doubts on the objectivity of the analysis.

The same complexities apply to PK similarity assessments, which use a similar hypothesis test framework at the outset. Thus, despite previous investigations (3–6) having shown that modeling gave similar results in BE analyses, we do not advocate replacing traditional (dense sampling) assessments by modeling in all situations. The traditional assessment better protects against Type I error by employing fewer assumptions and is less prone to controversies. In addition, using individual non-compartmental AUC and C_{max} values has the advantage of being model-free and is therefore robust to model misspecification. However, sparse sampling may occur for part of the clinical study data. In these situations where the traditional assessment is not possible, the modeling approach provides a rational alternative.

Our purpose is to draw attention to how to use modeling in situations where it is desirable to control the false-positive (Type I error) rate, for example, for claims submitted to regulatory authorities. Most of the above complexities relate to the issue of exploratory versus confirmatory analyses. In an exploratory analysis, the modeler typically examines the data first and then evaluates multiple models in order to find one that "best fits." Conclusions are typically drawn based on the best-fit model. The exploratory analyses are useful for models and predictions that are most likely, that is, for hypothesis generation. In contrast, in a confirmatory analysis such as BE assessment, the model is prespecified, very few (if any) model explorations are conducted, and Type I error rate is protected. Thus, in order to have confirmatory effects, the focus of model-based assessments needs to move toward confirmatory analyses.

On a more general level, the topic of what may constitute "confirmatory evidence" of a Phase 3 clinical study in NDA submissions has been heavily debated (9). One related issue is what level of evidence may be drawn from model-based analyses. We discuss this issue from the limited perspective of assessing PK similarity.

16.4   METHODOLOGY

We propose first that a population PK model be built. The average model parameters can be used to predict average PK endpoints, namely, AUC and C_{max}, in different formulations, and thus also the ratio of average AUC and C_{max} values. Then 90% confidence intervals for these ratios can be used for BE assessment. However, special attention is required for the model building process. Because the formulation effect was the central hypothesis for BE testing, every effort must be made to ensure that it is fully maintained during the modeling process. Consequently, the formulation effects should not be tested for the significance of their presence in the analysis. The rationale for this separate criterion for the formulation effect was similar to the situation of BE analyses, which aim to establish bounds on the formulation effect, and do not simply eliminate it even though it is often statistically not different from zero. The purpose was thus to make every effort to include the subject population effect, as long as practically possible.

In addition, model selection should be limited in order to protect against Type I error of BE testing. Therefore, we argue that a proper assessment framework should incorporate the following features:

1. The structural, covariate, and random effect models should be selected based on prior information and can be as complex as the data can be expected to support with high confidence.

2. Model explorations, if conducted at all, should be justified with suitable a priori power and should be consistent with the principle of BE assessment. For example, effects of formulation by covariate interaction should not be tested, for the same reason that, in standard BE analysis, treatment by subgroup interaction is typically not tested. This implies that the number of explorations should be kept very small, and preferably avoided in general.

3. Influence of the formulation factor, the key covariate of BE assessment, should be maintained on all structural model parameters (clearance, volume, etc.) and should not be eliminated based on the (in)significance of any statistical tests.

4. A BE assessment method that constructs 90% confidence intervals for AUC and C_{max} ratios of test/reference formulations must be prespecified.

Note that confidence interval construction for the C_{max} ratio represents a challenge because of the difficulty of formulating C_{max} as a model parameter. Bootstrap (10) allows this construction, though, because in each bootstrap run, the predicted C_{max} for the test and reference formulation, and thus their ratio, can be calculated from the population model parameters. The percentile bootstrap method then uses the 5% and 95% percentiles of the bootstrap runs to form the 90% confidence interval. Specifically, in each bootstrap run, a bootstrap data set can be generated where the subjects were resampled with replacement. Parameter estimates can be obtained for the bootstrap data set, and thus a ratio of AUC and C_{max}. Results of all bootstrap data sets can be assembled and the 5% and 95% percentiles used to construct the 90% bootstrap confidence intervals.

The computation of population average AUC and C_{max} deserves some elaboration. Common population models have interindividual variability of CL as lognormally distributed, in which case the population CL estimate is also the population geometric mean. Frequently, this can give rise to the average AUC through the relationship $Dose/CL$. Because AUC is commonly regarded as more important than C_{max}, it may be tempting to reparameterize the model using AUC instead of CL. Again, under the lognormal assumption, this will not change the parameter estimates. Thus, when intersubject variability of PK parameters is assumed lognormally distributed, the AUC for the average individual is the geometric mean of AUC values in the population.

The computation of C_{max} was more complex and dependent on all other model parameters. In most situations, no single model parameter would amount to the formulation (subject population) influence, unlike CL. Obtaining the average C_{max} is difficult, however, because the distribution of C_{max} cannot be directly calculated from the intersubject variability distribution of model parameters. Therefore, we suggest that the C_{max} of the typical individual in the population, computed from the

estimated population PK model parameters, be used instead. In theory, the C_{max} for the average individual could be somewhat different from the geometric mean of C_{max} in the population.

This framework addresses the difficulties mentioned in the previous section. Maintaining the formulation influence on all structural model parameters allows the model to adequately accommodate potential differences in the formulations addressed in BE assessment. Limiting the structural, covariate, or random effect model explorations allows proper interpretation of hypothesis test results. However, these choices can be controversial and are discussed in more detail next.

16.4.1   Structural Model

We first discuss the influence of structural model choice and the need of its pre-specification in BE assessment. On a rough scale, the *AUC* (and thus the ratio of test/reference) estimate should be relatively robust, because a model tends to represent the average concentration reasonably well. However, for C_{max} the opposite should hold. For an illustration, assume that the data arose from a two-compartment model with first-order absorption. Fitting a one-compartment model to the data would underestimate C_{max} and thus likely obscure any potential difference of C_{max} in the test and reference formulations. Therefore, using a less sophisticated model is likely to bias BE assessment (of C_{max}) toward equivalence.

The above argument could be extended to say that, because the reality is highly complex, any model will likely underestimate C_{max}. This may be true, but the extent should be taken in perspective. The traditional use of measured C_{max} also underestimates C_{max}, because it is unlikely that T_{max} is among the sampling times. A prerequisite of using modeling in any circumstance should be that the influence of potential model misspecification is limited, compared with alternative choices.

16.4.2   Covariate Model

In traditional BE assessment of standard crossover studies, adjusting for covariates can be shown to have no effect if the covariate values do not change over time. The use of a covariate model in population pharmacokinetics for BE assessment is analogous to adjusting for covariates in BE assessment of parallel studies. In standard BE assessment of parallel studies, adjusting for a covariate (e.g., gender) adds one term in the linear model, which costs only one degree of freedom but can potentially improve the error estimate and thus improve the power of BE assessment. With the use of population modeling, it is less clear where to add the gender effect, unless prior information is available. Adding gender influence to all model parameters not only reduces the degree of freedom, but also increases the complexity of parameter estimation in population modeling. Bootstrap could be particularly sensitive to model instability, because one needs a large number of model runs, and convergence could become problematic.

Covariate exploration could create potential trouble in interpreting BE assessment results, if any covariates were found to influence PK parameters. On the other hand, if existing literature evidenced some covariates that would affect the pharmacokinetics, it would be more difficult to argue for disregarding these covariates. In a certain sense, this represents a compromise between the exploratory and

confirmatory aspects of the analysis, depending on the information available. That is, the analysis could only afford to be more confirmatory when prior information is strong enough to justify those model assumptions are prespecified and not explored. The proposed approach allowed some explorations, although limiting them to only rational choices as suggested from previous experience. Those who might differ from this view could choose anywhere from one extreme to the other, perhaps depending on the purpose of the analysis. However, when model exploration is allowed, it is important that the analysis plan is prespecified and followed with rigor, in order to avoid analyst subjectivity. Even so, the statistical properties of the analysis results become worse as more exploration is allowed (7, 11). Intuitively, more model exploration would adversely affect the power of concluding the similarity of pharmacokinetics, because some degrees of freedom must be sacrificed for exploration. The issue would be much more apparent had an exploratory analysis arrived at some exotic model—for example, one that had ethnic origin influence absorption rate—that is both pharmacologically weak and unstable.

On the other hand, it is reasonable to argue that, because PK similarity assessment is the purpose of the investigation, the expectations and procedures should determine the philosophy and procedure of the analysis. PK modeling is merely a tool to achieve the goal. The relevant question is: "Is covariate exploration needed for the PK similarity assessment?" and not "Is covariate exploration important for building the PK model?" Thus, covariate exploration may cause more trouble than it's worth, viewed from a statistical perspective. One alternative might be to adjust for all of them without any pretesting. While this would better control Type I error, the power (or Type II error) could seriously suffer. A more interesting alternative is to assess the impact of including each covariate on PK similarity assessment. This may be suitable in certain situations. However, it also increases the complexity of the assessment, should any substantial impacts be found. Implementation may also be difficult when the computational burden is heavy. Finally, sensitivity analysis can be useful by comparing assessments with and without covariate exploration, had they resulted in different models.

16.4.3 Random Effect Models

Some random effect model parameters, especially the correlations among random effects, post little practical interest. For example, it is of little value to formally test whether the intercompartmental clearance varies by individual, or whether clearance and volume correlate. Most likely they do, and an "insignificance" usually only shows the lack of power to detect the effects. However, the real question is whether adjusting for such effects would improve the inference. Including such effects might underestimate model uncertainty. However, overfitting the model will dilute study power. We suggest that such terms not be included, especially in situations where power is lacking, pending further investigation.

16.4.4 Confidence Interval Construction

When confidence intervals are used as a form of hypothesis test, the Type I error is also affected by the correctness of the confidence interval bounds. Efron and Tibshirani (10) discussed the notion of correctness and accuracy of confidence

intervals. Common ways of constructing confidence intervals all involve varying degrees of accuracy, thus adversely affecting the Type I error. Obviously, a more accurate construction leads to better control of Type I error. We suggest bootstrap because it is conventionally believed as more accurate than using software-generated standard errors of parameter estimates. However, the precision of bootstrap depends on whether the numbers of subjects in the studies are large enough to be considered as the whole population. More experience on using bootstrap may need to be obtained; however, the emphasis here is on attempting to compute the confidence intervals more precisely and in a prespecified fashion as opposed to a post hoc analysis. If bootstrap is chosen as the analysis method, however, there are a few implementation issues:

1. *How to Bootstrap.* First, the number of subjects in a multistudy data set for the purposes presented needs to be kept constant to maintain the correct statistical interpretations of bootstrap, that is, correctly representing the underlying empirical distribution of the study populations. Second, the nonparametric bootstrap, as opposed to some other more parametric alternatives, was considered more suitable in order to minimize the dependence on having assumed the correct structural model.

2. *Number of Bootstrap Runs to Conduct.* This may depend on the desired precision and the type of data and thus is difficult to determine in general. Efron and Tibshirani (10) suggest that at least 2000 samples may be needed. In the authors' experience, 2000–5000 will generally suffice.

3. *How to Deal with Failed Bootstrap Runs.* Ette (12) provided valuable insights on this. It may be argued that failed runs correspond to data that were more likely to be in some way not typical of the population, that is, "out on tail." In estimating the 90% confidence interval, these types of samples are of particular interest, and runs need to be restarted until convergence is achieved. A counterargument to this could be that the successful runs were not paid similar attention; thus, the bias, if any, caused by ignoring failed runs may not be so severe. It could also be argued that when the software converges successfully but fails to produce standard error of parameter estimates, it presents only an algebraic problem unrelated to the quality of convergence; thus, such runs should be included. It may be tempting to conduct all these alternatives and examine the sensitivity, especially in a typical modeling scenario. However, doing so would violate the prespecified analysis plan and thus distracts from the confirmatory spirit. In hindsight, the problem can be alleviated by using multiple starting estimates for each run to increase the convergence rate. The increased computational burden may best be addressed using distributed computing (e.g., see `http://www.page-meeting.org/page/page2003/Sale.pdf`). Finally, an alternative approach that would avoid this problem is Markov chain Monte Carlo (MCMC) methods (13).

4. *Potential that a Different Interested Party* (e.g., a Regulatory Authority) *Could Obtain a Different Result by Using a Different Number of Samples or Different Inclusion Criteria.* In the authors' experience, this usually does not arise as an issue if the scientific rationale is presented clearly. However, when in doubt, it is always prudent to obtain an understanding from the interested party prior to conducting the analysis.

Future investigations would be necessary to determine the precision of bootstrap and how best to deal with failed runs. The reader is referred to Chapter 15 for more on implementation issues with the bootstrap.

16.4.5   Summary

Because of its many assumptions, a population model, especially with all the pre-specification demanded in this framework, is unlikely to be "true." However, one can argue that this framework exerts the influence of model misspecification primarily on study power. This is because a misspecified model would generally result in lower power although not larger Type I error. In addition, this approach maintains a more realistic confidence interval width instead of an overly optimistic (short) one. By "maximizing" the model as much as data can be expected to support, the impact on Type I error is minimized. Therefore, the hypothesis test is made as conservative as possible, and thus suitable for BE assessment.

This framework demonstrates a more general point: that is, how (or whether) to select the model should depend on the intended use of the model. In general, the relevant question for most applications of modeling is not whether the final model is right or wrong. As has been repeatedly quoted, "all models are wrong, but some are useful." The process of model selection should vary, depending on the goal. Ette et al. (14) discussed model appropriateness, and the use of bootstrap to estimate model prediction error, in a circumstance where the goal of modeling is prediction.

Due to the complex nature of population modeling, it can be difficult to foresee all potential outcomes and prespecify the corresponding strategies. If some unexpected outcomes that were not prespecified occur, it may be best to decide on a solution as close as possible to one that would have been prespecified. An alternative could be conducting several reasonable analyses and examining the robustness of the conclusions. This might be fine in some circumstances, however, the number of analyses could easily become impractical and arguably inflate Type I error.

16.5   APPLICATION EXAMPLE

We applied the methodology to a situation involving Phase 1–3 clinical study data. The work was conducted as part of a successful NDA submission, to respond to the US Food and Drug Administration's request that PK similarity be established in a situation where dense sampling was not available for all subjects.

16.5.1   Data

GW433908 is a phosphate ester prodrug of amprenavir (APV) and was being developed for treatment of HIV infection in adults and children. In four clinical studies, GW433908 was given alone and in conjunction with ritonavir to healthy subjects and HIV-infected subjects. PK samples were collected after single daily (QD) or twice daily (BID) dosing of at least 14 days, and tests based on serial trough sampling confirmed that steady-state amprenavir concentrations had been reached by day 14.

TABLE 16.1   Details of the Study Designs

Study Number (Population)	Dose	Number of Subjects[a]	Sampling Schedule[b] (h)	Assay	Number of Samples[c]
GW433908 Alone					
APV10013 (healthy)	GW433908 1400 mg BID × 14 days	12	0 (predose), 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 8, 10, and 12 h postdose at day 14	Advion	180
APV20001 (HIV-infected)	GW433908 1395 mg BID × 14 or 28 days	27	0 (predose), 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 6, 8, 10, and 12 h postdose at days 28 and 42	GSK	378
GW433908 + Ritonavir					
APV10009 (healthy)	GW433908 1395 mg + ritonavir 200 mg QD × 14 days	27	0 (predose), 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5, 8, 10, 12, 16, and 24[d] h postdose at day 14–15	GSK	375
APV30002 (HIV-infected)	GW433908 1395 mg + ritonavir 200 mg QD × 48 weeks	Day 28: 10 Week 4, 8, and 12: 35	0 (predose), 2, and 4 h postdose at day 28 0 (predose) at week 4, 8, 12	Advion	30 81
All combined		101			1044

[a] Number of potentially evaluable subjects.

[b] Study listing includes sampling schedule after repeat-dosing only.

[c] This number represents theoretical number samples that should be available for population PK analysis. Actual number may be lower due to various practical reasons (not recorded, assay interference, etc.).

[d] An additional sample was collected at 24 h for measurement of plasma unbound APV concentration.

Bioanalytical results were provided by two different sites using the same validated analytical techniques. Additional details of study designs are given in Table 16.1.

Several demographic variables were collected; however, weight, ethnic origin, and α-1-acid glycoprotein (AAG) were considered the more likely covariates to influence pharmacokinetics, based on previous data (15). A summary of demographics in actual data obtained for each study is given in Table 16.2.

Evaluation of APV pharmacokinetics following GW433908 administration to healthy subjects compared to HIV-infected subjects, both with and without ritonavir, was required. In the Phase 3 study (the last one in Table 16.1), only 3 samples

TABLE 16.2 Subject Demographics by Study[a]

Study Number	Number of Subjects	Ethnic Origin (W/B/O)	Gender (F/M)	Mean Age (SD) in Years	Mean Weight (SD) in kilograms	Mean AAG (SD)	Total Number of Samples
10009	27	19/6/2	4/23	40.3 (10.3)	77.0 (12.6)		405
10013	12	6/6/0	0/12	36.3 (10.2)	75.7 (10.5)	28.2 (6.9)	180
20001	22	13/6/3	4/18	33.1 (7.0)	71.1 (11.1)	18.0 (6.1)	305
30002	43	14/20/9	10/33	40.2 (10.5)	75.6 (15.2)		123
Total	104	52/38/14	18/86	38.3 (10.1)	75.0 (13.3)	21.6 (8.0)	1013

[a]W = white, B = black, O = others, F = female, M = male, SD = standard deviation.

at most were collected per subject. Individual AUC and C_{max} could not be reliably calculated by traditional means; therefore, assessing PK similarity based on PK parameters obtained from noncompartmental analysis of individual PK was not feasible.

16.5.2   Population Model Building

For the PK model structure, there was a fair amount of previous information, which suggested that oral one-compartment models might be sufficient to describe the PK profile, and there should be no need to consider models more complex than oral two-compartment models. For the potential covariates, ritonavir had been shown to strongly increase APV concentrations in previous studies. In addition, AAG might appear to influence clearance (16), and weight might influence volume and clearance, based on physiological considerations. In this analysis, some exploration was utilized, but the choices were limited to only those factors suggested from previous data. That is, oral one- and two-compartment models were considered, as well as the potential influences of AAG on clearance and weight on volume and clearance. In addition, the subject population (i.e., healthy or HIV-infected) was the key effect, similar to that of formulation effect in BE studies.

After the model-predicted ratios of average AUC and C_{max} in the populations were calculated, assessing PK similarity typically requires obtaining 90% confidence intervals for the ratios. The NONMEM-generated standard errors would be indicative but were known to be only approximate (17). Another estimate less dependent on distributional assumptions was obtained by using bootstrap to calculate the 90% confidence intervals.

The number of bootstrap samples generated would depend on the computational complexity, although at least 500 samples were to be attempted. When this plan was developed, the intention was that confidence intervals falling inside the (0.80, 1.25) range would demonstrate BE. A somewhat broader, although unspecified, range would show PK similarity. However, from a philosophical perspective, controversies exist as to the appropriateness of claiming BE among healthy and diseased populations. For a formal procedure of assessing PK similarity, a pre-specified range would be necessary to maintain the proper interpretation of the hypothesis test.

16.5.2.1   Analysis Plan

To minimize subjectivity in the model building process, the above was formalized in a prespecified analysis plan.

1. *Base Model Choice.* The choice was a steady-state one-compartment model with first-order absorption or a steady-state oral two-compartment model with first-order absorption. The disposition parameters were to be expressed in volume and clearance. Intersubject variability and residual error were also to be assessed. The best-fit model, using the software NONMEM, was to be the final base model. The criteria for accepting the NONMEM base model included (a) improved fitting of the diagnostic scatterplots (observed vs. predicted concentration, residual/weighted residual vs. predicted concentration

or time), (b) convergence of the minimization, (c) number of significant digits >3, (d) termination of the covariance step without warning messages, (e) absolute value of estimation correlation between model parameters <0.95, and (f) significant decrease in the objective function.

2. *Covariate Model Development.* Treatment (with and without ritonavir) would be included as a potential covariate on clearance. Influence of AAG on clearance and of weight on volume and clearance were also included as potential covariate relationships. Assay site was also included as a potential covariate for residual error.

With exception to the primary covariate (i.e., subject population), the covariate model development followed a step-forward–step-backward procedure. During the step-forward phase, the most promising covariate was added one-by-one, using the procedure similar to choosing the base model, where the decrease in the NONMEM objective function was required to be at least 3.84 for a single covariate ($p \leq 0.05$). During the step-backward phase, the least promising covariate was deleted one-by-one, using the procedure similar to choosing the base model, where the increase in the NONMEM objective function was required to be at least 6.63 for a single covariate ($p \leq 0.01$). This step-forward–step-backward procedure is common in PPK model building, the idea being that the choice of $p \leq 0.05$ in the step-forward phase allows all potential covariates to be considered, and that the choice of $p \leq 0.01$ in the backward elimination phase allows the inclusion of only sufficiently influential covariates.

Because the primary objective of this analysis was PK similarity assessment, the influence of subject population would first be assessed with a different criterion. The subject population effects would be parameterized as percent change, and not be tested by NONMEM objective function differences. However, to maintain numerical stability, a subject population effect would be dropped if the absolute value of the estimated effect were smaller than a nominal value of 0.01 (i.e., change of *CL*, *V*, K_a, etc. <1%). This was not expected to happen and was designed only to prevent the case where NONMEM might estimate these effects at infinitesimal values and consequently would not produce standard errors of parameter estimates.

After a model was built, the average *AUC* in the populations was to be calculated from the clearance parameter estimates (using $AUC = Dose/CL$), which correspond to clearances for the "typical" individuals in the populations, and thus the ratios of average *AUC* in the populations obtained. Under the usual population models, this gives the ratio of geometric means in the populations. Similarly, the ratio of average C_{max} in the populations could be calculated, although with more complex computations and a more subtle interpretation.

16.5.2.2 Model Building Result
The software NONMEM (17) with first-order conditional estimation (FOCE + INTER) was used throughout the analysis. The proportional error model was used for intraindividual variability, and interindividual variability was assumed lognormally distributed. Interindividual variabilities were assumed independent: that is, diagonal matrices for OMEGA were used throughout for model development. These details were not explicitly stated in the plan but were maintained as

if they were; that is, no explorations on the variability structures were considered during model building. The purpose was to avoid overspending degrees of freedom and resulting confidence intervals being too wide. The final structural model was a two-compartment model with first-order absorption. Study design did not allow the peripheral volume to be identified and it was fixed at an arbitrarily large value. One finer detail in the prespecified analysis plan was that study APV10009 (healthy volunteers) did not collect AAG; the covariate ethnic origin was to be used to impute AAG for this study based on the literature (15, 16). When the data became available, it turned out that AAG values were not collected in study APV30002 (HIV-infected subjects). Subsequently, AAG in two studies (APV10009 and APV30002) were imputed based on medians in ethnic origin groups, from APV10013 (healthy volunteers) and APV20001 (HIV-infected subjects), respectively. The analysis continued as planned and the results showed AAG values did not have any influence on APV population PK parameters. Since 50% of the AAG data were imputed, examination of ethnic origin as a potential covariate was conducted given its correlation with AAG (4, 16). Ethnic origin was "substituted" for AAG in the model since these data were available for all subjects and was also found to be insignificant. Therefore, the unexpected missing AAG data and imputation of AAG data and the examination of ethnic origin had little impact on the analysis.

Consistent with experience in other clinical studies, assay affected residual error, and treatment (with or without ritonavir) affected clearance. As stated above, no effect of weight, AAG, and ethnic origin was found. Absolute values of estimates for the effect of the primary covariate, subject population, on all model parameters were larger than 1% and thus retained in the model.

The final model was parameterized as follows:

```
Ka = θ₁*(1 + θ₇*POP)*exp(η₁)
CL = [θ₂*(1 – TMT) + θ₆*(1 – TMT)]*(1 + θ₈* POP)*exp(η₂)
V2 = θ₃*(1 + θ₉ * POP)*exp(η₃)
Q  = θ₄*(1 + θ₁₀ * POP)*exp(η₄)
```

The covariates were coded such that `TMT = 0` if without RTV, `TMT = 1` if with RTV, `POP = 0` if healthy, `POP = 1` if HIV-infected, and η_1, η_2, η_3, and η_4 were independently normally distributed. The residual error model took the form

```
Y = F + F*(1 + θ₅*ASSAY)*ε
```

where `ASSAY = 0` if GSK, `ASSAY = 1` if Advion. Parameter estimates are listed in Table 16.3.

The model assumes that the interindividual variabilities in the two populations are the same, which was implicit in the model building plan because no variability differences in the subject populations were to be tested. The plan was written based on the assumption that any such differences would have limited influence on the population average parameter estimates. This assumption is certainly debatable, however, in some sense similar to whether such differences should be assumed in formulations with standard BE assessments.

The observed concentrations (*DV*) are plotted versus the population predicted for the final model in Figure 16.1.

TABLE 16.3 Final Population PK Model

Model	Parameter (Unit)	Final Estimate	Standard Error
θ_1	Ka (h^{-1})	0.589	0.061
θ_2	CL (L/h)	91.8	7.1
θ_3	V2 (L)	164	11.1
θ_4	Q (L/h)	15.8	1.5
θ_5	ASSAY	0.437	0.107
θ_6	CL (L/h)	21.5	1.19
θ_7	POP on Ka	0.134	0.151
θ_8	POP on CL	−0.0372	0.0747
θ_9	POP on V2	−0.0764	0.0934
θ_{10}	POP on Q	0.342	0.245
η_1	—	0.199	0.051
η_2	—	0.094	0.016
η_3	—	0.089	0.031
η_4	—	0.207	0.052
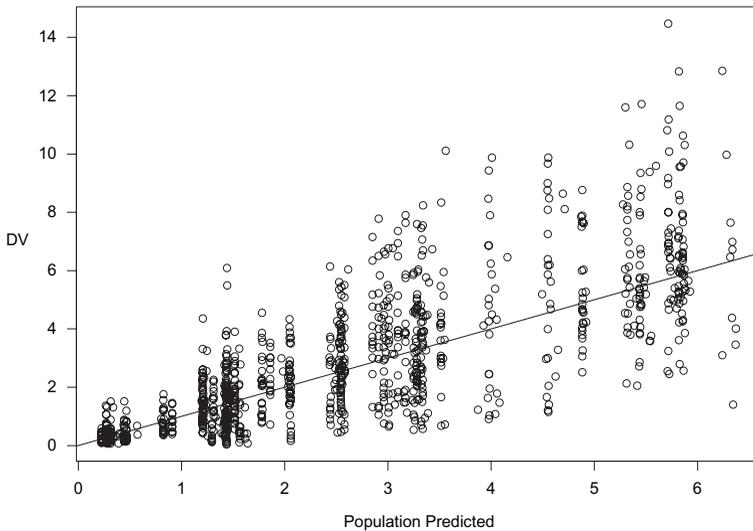
FIGURE 16.1 Final model population observed concentrations (DV) versus predicted concentrations. The predicted concentrations appear to take a series of fixed values corresponding to the number of observation time points.

Note that the final model was not to be further modified based on any diagnostics plots. Thus, it was comforting that Figure 16.1 suggested no specific misfits of the model. However, in general, many more diagnostic plots would be needed to assess goodness of fit of population models, if that were the primary focus. As an additional check of model performance, the model predicted AUC and C_{max} for the intensively sampled individuals were comparable with the noncompartmental analysis results.

16.5.3 PK Similarity Assessment

Bootstrapping was used to construct 90% confidence intervals for the estimates of the key variables, that is, the ratios of steady-state AUC and C_{max} of the typical individual for HIV-infected subjects over healthy subjects. The software S-Plus was used to generate the bootstrap data sets and to automate the NONMEM estimation. A total of 3000 bootstrap runs were conducted. This number was chosen due to computational complexity (1000 bootstrap runs required about 4 days using a 900 MHz personal computer). Among the 3000 runs conducted, 149 did not terminate successfully and were removed from further consideration. In retrospect, it would have been better to track the reason of convergence failure and attempt different starting values. Generally, a large proportion of missing bootstrap samples may bias the results, as described in Ette (12). For each of the remaining 2851 bootstrap runs, the parameter estimates for the typical individual were used to calculate steady-state AUC and C_{max} for the average healthy subject and HIV-infected subject population, and a ratio of AUC and C_{max} was obtained. In particular, AUC was calculated as *Dose*/*CL*. Calculation of C_{max} for two-compartment models in general would require simulation; however, for large peripheral volumes it can be computed as follows. Let the steady-state oral two-compartment model be parameterized as (`Ka`, `V2`, `CL`, `V3`, `Q`, τ). Standard textbooks give the concentration of steady-state oral two-compartment model by

```
c1*exp(-λ1*t) / (1 - exp(-λ1*τ)) + c2*exp(-λ2*t) / (1 - exp(-λ2*τ))
+ c3*exp(-Ka*t) / (1 - exp(-Ka*τ))
```

where

```
k10 = CL/V2, k12 = Q/V2, k21 = Q/V2
det = [(k10 + k12 + k21)² - 4*k21*k10)] ¹ᐟ²
λ1 = 0.5*( k10 + k12 + k21 + det), λ2 = 0.5*( k10 + k12 + k21 - det)
c1 = (Ka/V2)*(k21 - λ1)/ ( (λ2 - λ1)*(Ka - λ1))
c2 = (Ka/V2)*(k21 - λ2)/( (λ1- λ2)*(Ka - λ2))
c3 = (Ka/V2)*(k21 - Ka)/( (λ1- Ka)*( λ2 - Ka))
```

With large `V3`, one can show that this can be calculated as

```
d1*exp(-L1*t) / (1 - exp(-L1*τ)) + d2 - d1*exp(-Ka*t) / (1 - exp(-Ka*τ))
```

where `L1 = (CL + Q)/V2`. Taking the derivative of this with respect to `t` and setting it equal to 0 shows that it is maximized at

```
Tmax = 1/(Ka - L1) *log(R)
```

with

```
R = exp((Ka - L1)* Tmax) = Ka / (1 - exp(-Ka*τ)) / [L1 / (1 - exp(-L1*τ))].
```

The results of AUC and C_{max} in subgroups are given in Table 16.4.

TABLE 16.4   Bootstrap Results for *AUC* and *C*ₘₐₓ by Study, as Well as the Ratios for Subject Population

Variable	5% Percentile	Median	95% Percentile
AUC: healthy—GW433908	13.462	15.251	17.327
AUC: healthy—GW433908 + ritonavir	59.362	65.187	71.173
AUC: HIV-infected—GW433908	14.049	15.855	17.745
AUC: HIV-infected—GW433908 + ritonavir	60.518	67.460	74.763
Ratio of *AUC* (HIV-infected/healthy)	0.908	1.041	1.174
Unit of $AUC = \mathrm{h} * \mu\mathrm{g/mL}$			
C_{max}: healthy—GW433908	2.791	3.181	3.588
C_{max}: healthy—GW433908 + ritonavir	5.348	5.865	6.402
C_{max}: HIV-infected—GW433908	3.081	3.535	4.025
C_{max}: HIV-infected—GW433908 + ritonavir	5.791	6.391	7.063
Ratio of C_{max}, GW433908 only (HIV-infected/ healthy)	0.951	1.108	1.297
Ratio of C_{max}, GW433908 + ritonavir (HIV-infected/healthy)	0.956	1.088	1.244
Unit of $C_{max} = \mu\mathrm{g/mL}$			

Table 16.4 shows the 5% and 95% percentiles, which were used to form the 90% confidence intervals for BE assessment. The 90% confidence interval for the *AUC* ratio (HIV-infected vs. healthy) was (0.908, 1.175). This holds for both cases of with and without ritonavir, because *AUC* was determined by clearance, and the final model contained no interaction term of subject population and ritonavir on clearance, in a multiplicative sense. However, because C_{max} depends on all model parameters, the model predictions for the subject population effect on C_{max} are different for GW433908 only and GW433908 + ritonavir. Thus, the confidence intervals needed to be computed separately. The confidence interval for GW433908 alone was (0.951, 1.297), and that for GW433908 + ritonavir was (0.956, 1.244).

The confidence intervals were constructed from bootstrap runs that included 108 runs with failed covariance; that is, NONMEM was unable to generate standard errors of parameter estimates. Arguments could be made to include or exclude these runs in the analysis. Excluding these runs did not result in noticeable change of the results (i.e., changes on the confidence bounds <0.0005). Note also that a successful implementation of the NONMEM covariance step has no influence on the estimation of the geometric mean parameters. In retrospect, the analysis plan should prespecify whether such runs would be included, for the sake of rigorousness.

The clinical implication of this analysis is that the *AUC* and C_{max} in healthy and HIV-infected subjects were considered similar. This interpretation is less interesting from the methodological standpoint and is presented only for completeness.

16.5.4   Discussion

In hindsight, the analysis has room for improvement. The amount of model exploration, especially covariate searching, could be better planned. In that regard, it is fortunate that the final model in this particular application did not contain any

unexpected covariates influencing PK parameters. The only unexpected covariate in the final model was that the residual error model depended on the assay site. Thus, the covariate is unlikely to qualitatively affect the analysis results.

The covariate of with/without ritonavir may deserve more consideration. The question related to the central hypothesis test of PK similarity is: "Does the addition of ritonavir modify the conclusion about PK similarity?" From a statistical perspective, the ritonavir covariate may also deserve some special attention during model building, similar to the subject population covariate. However, practically, model stability (i.e., the replication stability of the final model form) decreases as more effects are estimated. In hindsight, it may be more appropriate to prespecify that the final model include an interaction term between subject population and the ritonavir covariate, and that ritonavir will influence the clearance only. This is in part because elevation of exposure of GW433908 when given with ritonavir prompted the inclusion of ritonavir in this assessment.

16.6   PHARMACODYNAMIC ENDPOINT BIOEQUIVALENCE

The need of pharmacodynamic (PD) BE assessment arose in the mid-1990s, and currently there is not yet a universally accepted analysis method with a corresponding study design. Here we discuss properties of some potential study designs and analysis methods based on the dose-scale approach, under the framework of two separate scenarios: broncodilation and broncoprovocation.

The dose-scale approach for PD BE assessment is based on a structural model described as follows. It assumes that the dose–response for the reference product follow the E_{max} model:

$$Response = E0 + Emax*Npuff /(ED50 + Npuff) \qquad (16.1)$$

The parameters are E0 (baseline response), Emax (maximum response), and ED50 (the dose that achieves 50% of maximum drug effect). Npuff is the number of puffs. In addition, it is assumed that the active ingredient in the test drug will be proportional to that of the reference drug, with a multiplication factor F. Under this assumption, the dose–response relationship for both the test and the reference drug can be written jointly as follows:

$$Response = E0 + Emax*Npuff*F^I /(ED50 + Npuff *F^I) \qquad (16.2)$$

where I = 0 or 1 indicates the reference or the test product, respectively. Parameters in the structural model are (F, E0, Emax, ED50). The parameter F is the relative bioavailability, which indicates bio(in)equivalence. (F = 1 would imply bioequivalence.) The rest of the parameters may be viewed as nuisance parameters. This model assumes that the test and reference products have the same E0 and Emax and differ only in their ED50 values.

Common study designs for assessing bioequivalence of MDIs are multiperiod crossover. Responses to the test and reference products of each subject are measured at baseline, at one or two doses of the test drug, and at two doses of the reference product. Due to the complexity of conducting MDI bioequivalence studies,

it is useful to have simple designs (i.e., as few doses as possible) and yet allow reasonable estimation of the relative bioavailability *F*. We consider two potential types of crossover designs:

1. *Limited Design.* Measurements are taken at the test baseline, test 1 puff, reference baseline, reference 1 puff, and one more reference at 2 puffs. Depending on the type of study, the extra reference dose may be at 4 puffs if it is felt that 2 puffs will not allow a good characterization of E_{max}.
2. *Balanced Design.* Add one more dose of the test drug to the Limited Design so that the test and reference are measured at the same doses.

We denote the doses as (`T0, T1, T2, R0, R1, R2`). For the limited design, `T2` is not available. The Balanced Design has been used in successful ANDAs. Intuitively, the Balanced Design would allow estimation of the E_{max} dose–response curve for both test and reference products. The Limited Design would be the minimal informative design in the sense that it would allow characterization of the E_{max} dose–response curve for the reference product. With that, the response at T1 would allow estimation of relative bioavailability *F*.

In practice, responses at the conducted doses may not plateau, so the E_{max} model may not be identified. In such a case, perhaps the only viable alternative is to consider the linear model:

```
Response = E0 + c*Npuff*Fᴵ
```
(16.3)

instead of Eq. (16.2). The relative bioavailabity *F* would be interpreted as a change in the slope. Because it may not be a priori clear as to when the linear model needs to be used, the potential impact of model misspecification needs to be considered.

In addition, it is plausible that the analysis above may be inaccurate because individual responses vary. One way to adjust for this, to some extent, might be to adjust for individual baseline values. Using individual change from baseline as response, the structural model in Eq. (16.2) becomes

```
Response = Emax* Npuff *Fᴵ/(ED50 + Npuff *Fᴵ
```
(16.4)

Intuitively, adjusting for baseline might be advantageous if the maximum changes from baseline are similar for all individuals.

The PD BE is assessed by 90% confidence interval for *F*, and the target intervals appeared to be case-specific, although larger than (0.80, 1.25). In principle, *F* and its confidence interval could be assessed with population models. Application of this approach to MDI bioequivalence studies have been reported (18–21). The reports did not show the exact forms of the models used. Thus, the robustness of the conclusions to the population model specification is unclear.

16.6.1   Method

Because sampling points are few in common PD BE designs, data allow estimating interindividual variability for very few parameters, usually no more than one or two. It is unclear as to how best to prespecify population models for PD BE

assessment, such as which parameters should have interindividual variability estimated. Thus, we consider nonlinear regression models, in which case the model specification, parameter estimation, and confidence interval construction are much less controversial.

We consider the above mentioned study designs together with a few corresponding analysis methods and explore their potential. Several analysis methods may be appealing, depending on the particular study design. For example, measurements of various individuals could be pooled. Alternatively, one could consider mixed effect models, accounting for intersubject variability in certain parameters. We discuss the impact of estimation and confidence interval construction methods separately.

16.6.1.1 Analysis Methods

Traditional bioequivalence assessment is based on a specific linear model with sequence and period effects (see Schuirmann (1)). Because the E_{max} curve is nonlinear, it is unclear as to how best to accommodate sequence and period effects. Given the desire of limiting study complexity, simplification in analysis method(s) is necessary to make them suitable in a potential regulatory setting. Thus, we will ignore period and sequence effects. We consider three potential methods to estimate the relative bioavailability F.

1. *Inverse Regression.* This method corresponds to the Limited Design. First, take geometric means of all measurements at each time of measurement and call these `M0`, `MR1`, `MR2`, `MT1`. Use (`M0`, `MR1`, `MR2`) in Eq. (16.1) to solve for parameters (`E0`, `ED50`, `Emax`). Then use the obtained parameters and MT1 in Eq. (16.2) to solve for `F`. With the balanced design, which gives response of the test product at 2 puffs, it is natural to generalize the method to the following.

2. *Means.* First, take geometric means of all measurements at each time of measurement. We thus have `MR0`, `MR1`, `MR2`, `MT0`, `MT1`, `MT2`. Then use Eq. (16.2) to jointly estimate all parameters under a chosen residual error model.

The geometric means step in the above two methods is appropriate when the error in response measurements (including model misspecification error) is lognormal distributed. The geometric means step can be replaced by arithmetic means when normally distributed errors are more appropriate. From a regulatory perspective, this choice should be made before data collection.

Naturally, one could think of using nonlinear regression to jointly estimate the parameters and F, without taking the means of the responses. This consideration leads to the following method.

3. *Naïve Pooled.* Use Eq. (16.2) to jointly estimate all parameters under a suitable error model assumption (additive, multiplicative, or constant *CV*). We considered three error models:

- Additive: `y = pred + `ε
- Lognormal: `y = pred*exp(`ε`)`
- Constant *CV*: `y = pred + pred*`ε

Here, `y` is the observed response, `pred` is the model predicted response, and ε is assumed normally distributed with mean 0 and variance σ^2.

Additionally, in a potential regulatory framework, a method of constructing 90% confidence interval for F must also be prespecified in order to control the Type I error. Several methods are possible. First, standard nonlinear regression packages provide asymptotic standard errors of parameter estimates. This readily leads to the construction of asymptotic confidence intervals, based on the normality assumption. Other methods include likelihood profile and bootstrapping. We considered these three methods briefly.

16.6.1.2   Evaluation of Design/Analysis Method: Theory and Simulations

We start by showing that, in principle, the means and the naïve pooled methods behave quite similarly. We shall call a data set balanced with respect to covariates if the numbers of observations are the same for any given combination of the covariates.

Lemma   With ordinary least-squares regression on balanced data, taking arithmetic means over the covariates does not affect the results of nonlinear regression estimation under additive (normal) error model assumption. Similarly, taking geometric means over the covariates does not affect the results of nonlinear regression estimation under multiplicative (lognormal) error model assumption.

The proof is shown by taking the means and showing that it does not change the maximizer of the likelihood. The proof is not given here. It follows directly from the lemma that the means methods theoretically perform the same as the corresponding naïve pooled methods. However, the means method is numerically more stable in our experience. This limited us to three estimation methods, namely, the naive pooled with the choice of additive, lognormal, or constant CV error models. We studied their relative performance by simulation.

Simulations were conducted under the following population model motivated by the broncodilation data from application Example 1 in Section 16.7.1. Equation (16.2) was used as the structural model, lognormal interindividual variability terms were placed on (E0, Emax, ED50), and the residual variability was assumed lognormally distributed. Estimation was carried out in NONMEM. Fixed effects were estimated as $(F, E0, Emax, ED50) = (0.79, 0.92, 60.3, 2.99)$. The variance–covariance matrix of interindividual variability terms was estimated as

```
var-cov(log(E0), log(Emax), log(ED50)) =
 1.9            -0.359     -2.23
-0.359           0.757      1.72
-2.23            1.72       6.76
```

In addition, the standard deviation of intraindividual variability was estimated as 0.32. The data did not allow reasonable estimation of potential interindividual variability of F. How much this model represents reality may be questionable, as no interindividual variability of F was allowed. However, there was reasonable agreement between the distributions of observed and model simulated data, and we hypothesized that this gives us a viable framework to evaluate the study design and analysis methods.

We used Monte Carlo simulation to gain insight of the performances of the designs and the analysis methods. As initial explorations, the simulations focused on the accuracy and precision of estimating relative bioavailability. For potential regulatory applications, it is important to study the effect of error model (mis)specification and the influence of structural model misspecification. Of particular interest are the following issues:

- How well does the means method perform when the models are correctly specified?
- How well does the means method perform when the error model is misspecified?
- How well does the inverse regression method perform when the models are correctly specified?
- How well does the means method perform when the structural model is misspecified?
- How well does the inverse regression method perform when the structural model is misspecified?
- Would it be advantageous to adjust for baseline?

To investigate these issues, we considered the following corresponding simulation scenarios for the use of study design/analysis method:

I. Geometric means
II. Naïve pooled, with constant *CV* error model
III. Inverse regression
IV. Geometric means, using a (misspecified) linear structural model (Eq. (16.3))
V. Inverse regression, using a (misspecified) linear structural model (Eq. (16.3))
VI. Geometric means adjusting for baseline, using structural model (Eq. (16.4))

For scenarios I–III and VI, the study design was chosen as using the original doses of 0, 1, and 4 puffs. Scenario IV and V aimed at studying the effect of model misspecification, that is, when the data do not allow accurate estimation of the E_{max} model. Because ED50 was 2.99 puffs in the data generation model, the study design using doses of 0, 1, and 2 puffs as chosen for scenarios IV and V.

In each simulation scenario, 1000 runs were simulated. In each run, a set of 40 subjects and their responses were generated based on the data generation model. This number of subjects was somewhat larger than normal trials conducted, based on the consideration that power may be lacking in available PD BE assessment situations. This difference was not expected to affect the relative performance of the methods. For each simulation run, the bias of estimating F was calculated, and the resulting 10%, 50%, and 90% percentiles of the distribution from the 1000 runs are given in Table 16.5.

In Table 16.5, the medians indicate the bias (or the lack thereof) of the study design/analysis methods, and the 10% and 90% percentiles relate to the precision. Results of scenarios I show that, when the correct structural model is assumed and the correct type of mean corresponding to the error model was taken, the means

TABLE 16.5 Performances of Various Methods Under Simulation

Scenario	Doses	Method	Structural Model	Error Model	Median	10%	90%
I	(0, 1, 4)	Means (geometric)	Correct	Correct	-0.01	-0.19	0.22
II	(0, 1, 4)	Naive pooled	Correct	Misspecified	0.03	-0.26	0.46
III	(0, 1, 4)	Inverse regression	Correct	NA	0.01	-0.18	0.25
IV	(0, 1, 2)	Means (geometric)	Misspecified	Correct	0.05	-0.07	0.16
V	(0, 1, 2)	Inverse regression	Misspecified	NA	0.08	0.24	0.45
VI	(0, 1, 4)	Means (geometric, adjusting for baseline)	Correct	Correct	0.01	-0.26	0.30

method was virtually unbiased. Results of scenario II show that, when the naïve pooled method misspecified the error model (constant *CV*), it did not cause much bias, but the precision became much worse. Results of scenarios III show that the inverse regression method performed virtually identically to the means method. Results of scenario IV and V show that, when the structural model was misspecified, the means method and the inverse regression method both appeared to be biased. However, the inverse regression was notably worse in both the magnitude and the spread of bias. Finally, comparing results of scenarios VI and I suggests that adjusting for baseline did not improve estimation precision.

The simulations therefore suggest that the means (or naïve pooled) method, without the intuitive adjusting for baseline, had the best estimation performance.

We also looked at confidence interval construction based on standard errors of parameter estimates. If the asymptotic 90% confidence intervals are truly accurate, then they should cover the true parameter 90% of the time. We examined the confidence interval coverages in simulation scenario I, shown in Figure 16.2.

The coverage probabilities were exact at 0 and 1, as they must be. However, as the probability gets closer to 0.5, the coverage probabilities became much wider than the "truth." Thus, the asymptotic confidence interval appeared to be unsuitable in a regulatory framework.

16.6.1.3 *Study Design and Estimation Recommendations*
For the two study designs under consideration, the previous section showed that the limited design is appropriate if the structural model is correctly specified. However, the design is sensitive to even slight model misspecification. The balanced design offers better protection. Thus, the choice of the design could be left at the sponsor's discretion, in case they have high confidence about the structural model and are willing to sacrifice robustness in order to reduce the number of treatment periods in the study.
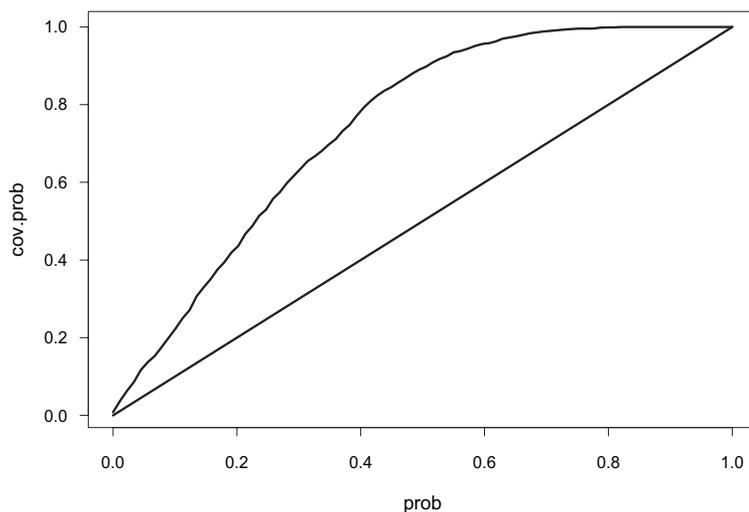
FIGURE 16.2 Coverage probabilities of asymptotic confidence intervals.

For the estimation methods, the naive pooled methods (or computationally more efficiently, the means methods) appeared reasonable when the error model is correctly specified. From the perspective of protecting against Type I error, the error model could be prespecified, based on prior experience of the type of studies (i.e., broncoprovocation or broncodilation).

We briefly explore later the potential of using likelihood profile and bootstrap as confidence interval construction methods through application examples with broncodilaion and broncoprovocation data.

16.7   APPLICATION EXAMPLES

We considered two data sets arising from ANDAs (data modified), which will be used to evaluate potential study design and analysis methods.

16.7.1   Example 1: Broncodilation Study

Twenty-four subjects completed an 8-period balanced crossover study. For each subject, pharmacodynamic responses are taken at the test baseline, test 1 puff, test 4 puffs, reference baseline, reference 1 puff, and reference 4 puffs. A population model fit (Section 16.6.1.2) was conducted to describe the data. Responses at baseline were similar for both formulations, although they appeared to differ somewhat at 1 puff and 4 puffs. Overall, the profiles of the two formulations were judged as similar.

The data clearly preferred the lognormal distribution to normal. Thus, the geometric means method was applied, using Eq. (16.2) as the structural model, which gave parameter estimates `(F, E0, Emax, ED50) = (1.06, 0.903, 30.5, 2.33)`.

For confidence interval of F, we applied the percentile bootstrap method. With 2000 runs, the 90% confidence interval was (0.74, 2.18). This appeared to be a reasonable quantification of the variability with the data.

16.7.2   Example 2: Another Broncodilation Study

Sixty-six subjects completed a 6-period balanced crossover study. For each subject, pharmacodynamic responses are taken at the test baseline, test 1 puff, test 2 puffs, reference baseline, reference 1 puff, and reference 2 puffs for a total of 66 subjects. Population modeling was attempted to describe the data, which has an AUC type of endpoint. However, it was difficult to estimate interindividual variabilities for all three E_{max} model parameters. Nevertheless, responses for the test and reference appeared sufficiently similar at baseline and both 1 puff and 2 puffs.

The data did not show any preference of the lognormal distribution over the normal. Thus, the arithmetic means method was applied, using Eq. (16.2) as the structural model, which gave parameter estimates `(F,E0,Emax,ED50)= (1.14, 692, 205, 0.566)`.

For confidence interval of F, we applied the percentile bootstrap method. With 2000 runs, the 90% confidence interval was (0.94, 1.44). Again, this appeared to be quite reasonable quantification of the variability with the data.

We also attempted the likelihood profile method to construct the 90% confidence interval. However, the likelihood profile of *F* turned out to be extremely flat in this case. As a result, the 90% confidence interval included (0.1, 10). This was considered unreasonable, given the similarity between the two formulations. Thus, the likelihood profile method did not seem suitable for confidence interval construction.

Programming codes written in NONMEM and S-Plus for the application examples are provided in the appendix (see Appendix 16.1).

16.8   SUMMARY

We were searching for study designs and the corresponding estimation and confidence interval construction methods suitable in a potential regulatory framework. The investigation is still sketchy and more complete assessment is necessary. In addition, the number of clinical trial data is few and more experience is certainly necessary. However, the findings seem reasonable and serve to pare down the potential choices. Based on the various experience in simulation and application examples, we recommend the balanced design, using the appropriate means method for estimation. The use of geometric versus arithmetic means should be chosen based on prior information on whether the data would be normally or lognormally distributed.

For confidence interval construction, the bootstrap performed the best among the possible methods considered. Thus, we recommend bootstrap as the method of choice.

REFERENCES

1. D. J. Schuirmann, A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* **15**(6):657–680 (1987).

2. *Interim Guidance for Documentation of in Vivo Bioequivalence of Albuterol Inhalation Aerosols (Metered Dose Inhalers).* Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research, FDA, January 27, 1994.

3. N. Kaniwa, N. Aoyagi, H. Ogata, and M. Ishii, Application of the NONMEM method to evaluation of the bioavailability of drug products. *J Pharm Sci* **79**(12):1116–1120 (1990).

4. G. A. Maier, G. G. Lockwood, J. A. Oppermann, et al., Characterization of the highly variable bioavailability of tiludronate in normal volunteers using population pharmacokinetic methodologies. *Eur J Drug Metab Pharmacokinet* **24**(3):249–254 (1999).

5. H. S. Pentikis, J. D. Henderson, N. L. Tran, and T. M. Ludden, Bioequivalence: individual and population compartmental modeling compared to the noncompartmental approach. *Pharm Res* **13**(7):1116–1121 (1996).

6. D. M. Reith, W. D. Hooper, J. Parke, and B. Charles, Population pharmacokinetic modeling of steady state carbamazepine clearance in children, adolescents, and adults. *J Pharmacokinet Biopharm* **28**(1):79–92 (2001).

7. S. Senn, *Statistical Issues in Drug Development.* Wiley, Hoboken, NJ, 1997.

8. J. J. Faraway, On the cost of data analysis. *J Comput Graphical Stat* **1**:213–229 (1992).

9. C. C. Peck and J. Wechsler, Report of a workshop on confirmatory evidence to support a single clinical trial as a basis for new drug approval. *Drug Inf J* **36**:517–534 (2002).

10. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap.* Chapman & Hall, New York, 1993.

11. F. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer-Verlag, New York, 2001.

12. E. I. Ette, Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* **37**:486–495 (1997).

13. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London, 1996.

14. E. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M. Liu, and E. V. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).

15. J. A. Johnson, Influence of race or ethnicity on pharmacokinetics of drugs. *J Pharm Sci* **86**:1328–1333 (1997).

16. B. M. Sadler, C. Gillotin, Y. Lou, and D. S. Stein, In vivo effect of alpha-1-acid glycoprotein on pharmacokinetics of amprenavir, a human immunodeficiency virus protease inhibitor. *Antimicrob Agents Chemother* **45**:852–856 (2001).

17. S. L. Beal and L. B. Sheiner, *NONMEM User's Guide.* NONMEM Project Group, UCSF, 1992.

18. P. Creticos, W. Adams, B. Petty, L. Lewis, and G. Singh, A methacholine challenge dose–response study for development of a pharmacodynamic bioequivalence methodology for albuterol metered-dose inhalers. *J Allergy Clin Immunol* **110**:713–720 (2002).

19. H. Leon-Molina, F. Flores-Murrieta, and R. Chapela, Assessment of comparative bioequivalence of two metered-dose inhaler formulations of salbutamol—measuring bronchodilatory effect in patients with asthma. *Clin Drug Invest* **22**(7):435–441 (2002).

20. K. Parameswaran, Concepts of establishing clinical bioequivalence of chlorofluorocarbon and hydrofluoroalkane β-agonists. *J Allergy Clin Immunol* **104**(6):S243–S245 (1999).

21. G. Singh, W. Adams, L. Lesko, et al., Development of in vivo bioequivalence methodology for dermatologic corticosteroids based on pharmacodynamic modeling. *Clin Pharmacol Ther* **66**(4):346–357 (1999).

APPENDIX 16.1

Programming codes written in NONMEM and S-Plus for the application examples are given below.

Programming Codes for Section 16.5

Data have been modified from the real data, for confidentiality reasons.

NONMEM Code

```
$PROB Application Example 1 - Final Population Model
$DATA boot.dat IGNORE=#
$INPUT PROT ID TIME EVID DV MDV AMT II SS TMT POP WT ASSA
$SUB ADVAN4 TRANS4
$PK
```

```
 KA = THETA(1)*(1 + THETA(7)*POP)*EXP(ETA(1))
 CLTYP = (THETA(2)*(1-TMT) + THETA(6)*TMT)*(1 + THETA(8)*POP)
 CL = CLTYP*EXP(ETA(2))
 V2 = THETA(3)*(1 + THETA(9)*POP)*EXP(ETA(3))
 Q = THETA(4)*(1 + THETA(10)*POP)*EXP(ETA(4))
 V3 = 8000
 S2 = V2
$ERROR
 IPRE = F
 W = F
IRES = DV - IPRE
Y = IPRE*(1 + (1 + THETA(5)*ASSA)*EPS(1))

$THETA (0, 0.591) ; KA
 (0, 91.9) ; CL
 (0, 164) ; V2
 (0, 15.8) ; Q
 (-0.5, 0.438) ; ASSA
 (0, 21.5) ; CL for + RTV
 (-1, 0.01) ; POP effect on KA
 (-1, -0.037) ; POP effect on CL
 (-1, -0.078) ; POP effect on V2
 (-1, 0.3) ; POP effect on Q

$OMEGA 0.2 0.094 0.09 0.21
$SIGMA 0.1

$EST MAX=5000 NOABORT METHOD=1 INTER
$COV
```

S-Plus Code

```
### NONMEM plots
NMtable_read.table("c:/temp/ NMtable.txt", header=T, skip=1)
NMtable_NMtable[NMtable$MDV==0,]
plot(NMtable$PRED, NMtable$DV); abline(0,1)
plot(NMtable$IPRE, NMtable$DV); abline(0,1)
plot(NMtable$PRED, NMtable$RES); abline(0,0)
plot(NMtable$PRED, NMtable$WRES); abline(0,0)
  lines(lowess(NMtable$PRED,NMtable$WRES), lty=3)
plot(NMtable$IPRE, NMtable$IRES); abline(0,0)
plot(NMtable$IPRE, NMtable$IWRE); abline(0,0)
  lines(lowess(NMtable$IPRE,NMtable$IWRE), lty=3)

### ETA
ETAcov_ read.table("c:/temp/ ETAcov.dat", header=T, skip=1)
ETAcov$TMT_as.factor(as.character(ETAcov$TMT))
bwplot(TMT~ETA1, ETAcov, main="ETA1 vs TMT")
bwplot(TMT~ETA2, ETAcov, main="ETA2 vs TMT")
```

```
bwplot(TMT~ETA3, ETAcov, main="ETA3 vs TMT")
bwplot(TMT~ETA4, ETAcov, main="ETA4 vs TMT")


### compute NONMEM bootstrap results
lim.SS.oral.cp2_function(param, time) {
  ka_param[1]; CL_param[2]; V1_param[3]; CLd_param[4]; tau_param[5]
  lambda1_(CL + CLd)/V1

  c1_(ka/V1)/(ka - lambda1)
  const_1/(lambda1*V1*tau) *CLd/CL

  c1*exp(-lambda1*time)/(1 - exp(-lambda1*tau)) +
  const -
  c1*exp(-ka*time)/(1 - exp(-ka*tau))
}
Tmax.lim.SS.oral.cp2_function(param) {
  ka_param[1]; CL_param[2]; V1_param[3]; CLd_param[4]; tau_param[5]
  lambda1_(CL + CLd )/V1
  right.hand_ka/(1 - exp(-ka*tau)) / (lambda1/(1 - exp(-lambda1*tau)))
  1/(ka - lambda1) *log(right.hand)
}
Cmax.lim.SS.oral.cp2_function(param) {
  tmax_Tmax.lim.SS.oral.cp2(param)
  lim.SS.oral.cp2(param, tmax)
}
## calculate parameters CL, V, etc.
## from bootstraped results NM.est.all


# AUC
boot.AUC.healthy.BID.CI_quantile(1400/NM.est.all[,2], c(0.05, 0.5,
0.95), na.rm=T)
boot.AUC.healthy.RTV.BID.CI_quantile(1395/NM.est.all[,6], c(0.05,
0.5, 0.95), na.rm=T)
boot.AUC.patient.BID.CI_quantile(1400/(NM.est.all[,2]*(1 + NM.est.
all[,8])),
  c(0.05, 0.5, 0.95), na.rm=T)
boot.AUC.patient.RTV.BID.CI_quantile(1395/(NM.est.all[,6]*(1 + NM.
est.all[,8])),
  c(0.05, 0.5, 0.95), na.rm=T)
boot.AUC.CI_quantile(1/(1 + NM.est.all[,8]), c(0.05, 0.5, 0.95),
na.rm=T)


### Cmax
# healthy
NM.est.healthy.BID_NM.est.all[,1:4]
NM.est.healthy.BID_cbind(NM.est.healthy.BID, rep(12, nrow(NM.est.
healthy.BID)))
```

```
# QD = + RTV
NM.est.healthy.QD_NM.est.all[,c(1,6,3,4)]
NM.est.healthy.QD_cbind(NM.est.healthy.QD, rep(24, nrow(NM.est.
healthy.QD)))
# patients
NM.est.patient.BID_NM.est.all[,1:4]
for (i in 1:4) {
  NM.est.patient.BID[,i]_NM.est.patient.BID[,i]*(1 + NM.est.
all[,i+6])
}
NM.est.patient.BID_cbind(NM.est.patient.BID, rep(12, nrow(NM.est.
patient.BID)))


NM.est.patient.QD_NM.est.all[,c(1,6,3,4)]
for (i in 1:4) {

  NM.est.patient.QD[,i]_NM.est.patient.QD[,i]*(1 + NM.est.all[,i+6])
}
NM.est.patient.QD_cbind(NM.est.patient.QD, rep(24, nrow(NM.est.
patient.QD)))


# Cmax
boot.Cmax.healthy.BID_1400*apply(NM.est.healthy.BID, 1, Cmax.lim.
SS.oral.cp2)
boot.Cmax.patient.BID_1395*apply(NM.est.patient.BID, 1, Cmax.lim.
SS.oral.cp2)
boot.Cmax.healthy.QD_1395*apply(NM.est.healthy.QD, 1, Cmax.lim.
SS.oral.cp2)
boot.Cmax.patient.QD_1395*apply(NM.est.patient.QD, 1, Cmax.lim.
SS.oral.cp2)


# Cmax ratios
boot.Cmax.BID_boot.Cmax.patient.BID/boot.Cmax.healthy.BID
boot.Cmax.QD_boot.Cmax.patient.QD/boot.Cmax.healthy.QD


# C.I.
boot.Cmax.healthy.BID.CI_quantile(boot.Cmax.healthy.BID, c(0.05,
0.5, 0.95), na.rm=T)
boot.Cmax.patient.BID.CI_quantile(boot.Cmax.patient.BID, c(0.05,
0.5, 0.95), na.rm=T)
boot.Cmax.healthy.QD.CI_quantile(boot.Cmax.healthy.QD, c(0.05,
0.5, 0.95), na.rm=T)
boot.Cmax.patient.QD.CI_quantile(boot.Cmax.patient.QD, c(0.05,
0.5, 0.95), na.rm=T)
boot.Cmax.BID.CI_quantile(boot.Cmax.BID, c(0.05, 0.5, 0.95),
na.rm=T)
```

```
boot.Cmax.QD.CI_quantile(boot.Cmax.QD, c(0.05, 0.5, 0.95),
na.rm=T)

# show
boot.AUC.healthy.BID.CI
boot.AUC.healthy.RTV.BID.CI
boot.AUC.patient.BID.CI
boot.AUC.patient.RTV.BID.CI
boot.AUC.CI
boot.Cmax.healthy.BID.CI
boot.Cmax.patient.BID.CI
boot.Cmax.healthy.QD.CI
boot.Cmax.patient.QD.CI

boot.Cmax.BID.CI
boot.Cmax.QD.CI
```

Programming Codes for Section 16.7

For confidentiality reasons, data have been simulated based on original data. The corresponding S-Plus codes are provided below.

Application Example 1 (Section 16.7.1)

```
# Data Generation
app1.dat <- data.frame(subj=rep(seq(1,24), each=6),
  puffs=rep(c(0,1,4), 48), trt=rep(c(0,0,0,1,1,1), 24))
rb <- 1.061698; e0 <- 0.9033722; emax <-  30.50206; ed50 <- 2.327858
app1.dat$resp <-
  (e0 + (emax - e0)*app1.dat$puffs*rb^app1.dat$trt/
    (ed50 + app1.dat$puffs*rb^app1.dat$trt))*
  exp(rnorm(n=144, sd=0.32))

# Dose-scale approach using geometric means
app1.geomeans.fun <- function(data) {
  geomean.dat <- aggregate(data$resp,
    list(trt=data$trt, puffs=data$puffs),
    function(x) prod(x)^(1/length(x)))
  for (i in 1:2) geomean.dat[,i] <-
    as.numeric(as.character(geomean.dat[,i]))
  coef(nls(x ~ e0 + (emax - e0)*puffs*rb^trt/(ed50 + puffs*rb^trt),
    start = list(rb=1.1, e0=0.9, emax=31, ed50=2.3),
    data=geomean.dat))
}
app1.geomeans.fun(app1.dat)

# bootstrap
app1.flat.mat_matrix(app1.dat$resp, ncol=6, byrow=T)
```

```
app1.geomeans2.fun <- function(mat) {
  data <- app1.dat
  data$resp <- as.vector(t(mat))
  geomean.dat <- aggregate(data$resp,
    list(trt=data$trt, puffs=data$puffs),
    function(x) prod(x)^(1/length(x)))
  for (i in 1:2) geomean.dat[,i] <-
    as.numeric(as.character(geomean.dat[,i]))
  coef(nls(x ~ e0 + (emax - e0)*puffs*rb^trt/(ed50 + puffs*rb^trt),
    start = list(rb=1.1, e0=0.9, emax=31, ed50=2.3),
    data=geomean.dat))
}

#app1.geomeans2.fun(app1.flat.mat)
boot.nls <- bootstrap(app1.flat.mat,
  app1.geomeans2.fun, B=2000)
#names(boot.nls)
quantile(boot.nls$replicates[,1], c(0.05, 0.5, 0.95))
```

Application Example 2 (Section 16.7.2)

```
# Data Generation
app2.dat <- data.frame(subj=rep(seq(1,66), each=6),
  puffs=rep(c(0,1,2), 132), trt=rep(c(0,0,0,1,1,1), 66))
rb <- 1.14088; e0 <- 691.9227; emax <- 896.7382;    ed50 <- 0.566
app2.dat$resp <-
  e0 + (emax - e0)*app2.dat$puffs*rb^app2.dat$trt/
    (ed50 + app2.dat$puffs*rb^app2.dat$trt) +
  rnorm(n=396, sd=197)

# Dose-scale approach using arithmetic means
app2.means.fun <- function(data) {
  mean.dat <- aggregate(data$resp,
    list(trt=data$trt, puffs=data$puffs), mean)
  for (i in 1:2) mean.dat[,i] <-
    as.numeric(as.character(mean.dat[,i]))
  nls(x ~ e0 + (emax - e0)*puffs*rb^trt/(ed50 + puffs*rb^trt),
    start = list(rb=1.1, e0=691, emax=896, ed50=0.566),
    data=mean.dat)
}
app2.nls_app2.means.fun(app2.dat)

# bootstrap
app2.flat.mat_matrix(app2.dat$resp, ncol=6, byrow=T)
app2.means2.fun <- function(mat) {
  data <- app2.dat
  data$resp <- as.vector(t(mat))
  mean.dat <- aggregate(data$resp,
```

```
   list(trt=data$trt, puffs=data$puffs), mean)
 for (i in 1:2) mean.dat[,i] <-
   as.numeric(as.character(mean.dat[,i]))
 coef(nls(x ~ e0 + (emax - e0)*puffs*rb^trt/(ed50 + puffs*rb^trt),
   start = list(rb=1.1, e0=691, emax=896, ed50=0.566),
   data=mean.dat))
}
#app2.means2.fun(app2.flat.mat)
boot.nls <- bootstrap(app2.flat.mat,
 app2.means2.fun, B=2000)
#names(boot.nls)
quantile(boot.nls$replicates[,1], c(0.05, 0.5, 0.95))

## examine likelihood profile
#F=0.1
F01.dat_ app2.dat
param(F01.dat, "e0")_ 691.9227;
param(F01.dat, "em0")_204.8155; param(F01.dat, "ed50")_0.566
F01.nls_nls(resp ~ e0 + em0*puffs*(0.1*trt + (1 - trt))
/(ed50 + puffs*(0.1*trt + (1 - trt))), data=F01.dat)

length(F01.nls$residuals)*log(sum(F01.nls$residuals^2))
pchisq(
length(F01.nls$residuals)*log(sum(F01.nls$residuals^2)) -
length(app2.nls$residuals)*log(sum(app2.nls$residuals^2))
,1)

#F=1000
F1000.dat_ app2.dat
param(F1000.dat, "e0")_ 691.9227;
param(F1000.dat, «em0»)_10004.8155; param(F1000.dat, «ed50»)_0.566
F1000.nls_nls(resp ~ e0 + em0*puffs*(1000*trt + (1 - trt))
/(ed50 + puffs*(1000*trt + (1 - trt))), data=F1000.dat)
length(F1000.nls$residuals)*log(sum(F1000.nls$residuals^2))
pchisq(
length(F1000.nls$residuals)*log(sum(F1000.nls$residuals^2)) -
length(app2.nls$residuals)*log(sum(app2.nls$residuals^2))
,1)
```

PHARMACOKINETICS/ PHARMACODYNAMICS RELATIONSHIP: BIOMARKERS AND PHARMACOGENOMICS, PK/PD MODELS FOR CONTINUOUS DATA, AND PK/PD MODELS FOR OUTCOMES DATA

Biomarkers in Drug Development and Pharmacometric Modeling

PAUL J. WILLIAMS and ENE I. ETTE

17.1 INTRODUCTION

During the last two decades the direct cost of drug development has continued to escalate at two and one-half times the rate of inflation. The cost of introducing a drug to the market was $802 million (US) in 2000 compared to $237 million in 1987 (1). As an indirect cost it takes 7–12 years for a drug to move through development to the final FDA approval (1). Several factors have influenced the escalation in the cost of drug development including an increased cost of executing clinical trials and more rigorous approval standards. Regulatory standards are not likely to become less rigorous; therefore, one must look elsewhere to improve the process. Extended use and novel applications of biomarkers may improve the drug development process by aiding in the construction of powerful and efficient clinical programs.

The Biomarkers Definitions Working Group (BDWG) has stated: "One approach to the achievement of more expeditious and informative therapeutic research is the use of clinical measurement tools to determine disease progression and the effects of interventions (drugs, surgery, and vaccines). . . . Another approach is the use of a wide array of analytical tools to assess biological parameters, which are referred to as biomarkers" (2). The discovery and utilization of biomarkers has several ways in which they could bring efficiencies and provide insight into the drug development process and patient care. Biomarkers can identify patients at risk for a disease, predict patient response, predict the occurrence of toxicity, and predict exposure to drug. Given these uses they can also provide a basis for selecting lead compounds for development and contribute knowledge about clinical pharmacology. Therefore, biomarkers have the potential to be one of the pivotal factors to lead drug development from drug target discovery to preclinical development to clinical development to regulatory approval and labeling information, by way of pharmacokinetic (PK)—pharmacodynamic (PD)—outcomes modeling with clinical trial simulations.

The flow of the chapter begins with establishing a common language for addressing the biomarker issue starting with natural history markers (Type 0 markers) and progressing to surrogate endpoints and clinical endpoints. We describe how a natural history marker can mature into a surrogate endpoint and how advances in technology are providing more and improved biomarkers. Then we go on to a description of how pharmacometric (PM) modeling interacts with biomarker technology and finish with an example of how the integration of PM modeling can result in a decrease in development time, lower development cost, and robust trial structures. First, we need a common vocabulary to organize our thinking.

17.2   VOCABULARY

The BDWG has provided a common vocabulary so that a common ground may be occupied when the topic of biomarkers is addressed. The following are the definitions for the current chapter.

17.2.1   Biological Marker (Biomarker)

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, a pathogenic process, or pharmacologic responses to a therapeutic intervention.

17.2.2   Natural History or Type 0 Markers

Natural history or Type 0 markers are those markers that measure disease predisposition, severity, or outcome, and reflect underlying pathogenetic mechanisms (3). Type 0 markers predict clinical outcome independent of treatment. They are often used to define inclusion or exclusion criteria for patients considered for enrollment into clinical trials, for stratifying these patient populations (because they indicate disease stage), or as milestones of disease progression for monitoring patients. Type 0 markers have a biological plausibility that is foundational for further development as biological activity markers.

17.2.3   Drug Activity Marker, Biological Activity Marker, or Type I Marker

A drug activity marker or Type I marker is a marker that reflects a response to therapy or drug treatment (3). The degree and magnitude of the response of the marker to drug therapy should correlate with the potency of the therapeutic agent. They are used to demonstrate proof of concept, to establish dose regimens, and for optimizing combination therapies. The degree of response can be used to determine optimal dosing strategies and to indicate whether combined therapy is more active than a single treatment. They are often pharmacodynamic response markers where the magnitude of the change defines the potency of the drug.

17.2.4   Surrogate Endpoint or Type II Marker

A surrogate endpoint or Type II marker is a biomarker that is intended to substitute for a clinical endpoint (2, 3). The US Food and Drug Administration has noted the

acceptance of surrogate endpoints to grant accelerated marketing for approving therapeutic agents. These are covered in Title 21 code of Federal Regulations Part 314 Section 510, subpart H which states:

> FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity. Approval under this section will be subject to the requirement that the applicant study the drug further, to verify and describe its clinical benefit, where there is uncertainty as to the relation of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome.

The ultimate stage in the development of a biomarker is when in the context of an effective therapeutic regimen, the relationship between early change in the biomarker and ultimate clinical outcome allows the use of the biomarker as a substitute for the clinical endpoint. A surrogate endpoint is expected to predict clinical benefit, harm, lack of benefit, or lack of harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. A surrogate endpoint can be either a single marker or a composite of several markers, which fully accounts for the efficacy of the agent being tested. Surrogate endpoints are a subset of biomarkers. All surrogate endpoints are biomarkers. However, few biomarkers will ever become surrogate endpoints. The term surrogate marker should be avoided.

17.2.5 Clinical Endpoint

A clinical endpoint is a characteristic or variable that reflects how a patients feels, functions, or survives. It is a distinct measurement of or analysis of disease characteristics observed in a study or a clinical trial that reflect the effect of a therapeutic intervention. Clinical endpoints are the most credible characteristics used in the assessment of the benefits and risks of a therapeutic intervention in randomized clinical trials.

17.3 BIOMARKER VALIDATION

Validation of a biomarker begins with the description of the pathogenesis of a disease and culminates when it is determined that the biomarker is applicable to clinical trials. This validation process follows a stepwise process depending on the stage of drug development.

17.3.1 Validation of Type 0 Markers or Natural History Markers

Natural history makers are validated when at baseline a strong relationship between the level of the marker and the ultimate clinical outcome of the disease has been established. These markers are often established in the placebo arms of early drug studies. Here the establishment of a sample repository can permit poststudy analysis to establish the veracity of a Type 0 marker.

17.3.2   Validation of Biological Activity Type I Markers

Type I markers are most often validated in Phase 1 or 2 clinical trials, where it is often demonstrated that the therapeutic intervention favorably changes the marker. Here, often dose-related effects can be established in studies where dose escalation is executed. The best place to establish a Type I marker is in Phase 1, especially when there is a placebo control group so that the natural history of the marker can be followed in the placebo group.

17.3.3   Validation of Surrogate Endpoints or Type II Markers

17.3.3.1   *General Criteria for Surrogacy*

A marker validated as Type 0 or Type I may next be validated as a Type II marker or surrogate endpoint. The validation of surrogate endpoints is best done with data from Phase 2 or 3 studies, where dropout rates have been low, the treatment has continued unchanged over the duration of the study, the biomarker was measured early in treatment, patients have been followed for a sufficiently long time, and a difference between control and treatment (especially for placebo-controlled trials) was demonstrated. Note in the case of no difference placebo-controlled trials the biomarker then would be a Type 0 marker. It is desirable to have this type of data from several studies to establish the validity of the marker across studies, drugs, and various patient populations.

When well defined clinical endpoints such as survival, end organ damage, or recurrence of cancer are employed for the establishment of efficacy, long periods may be required to observe these clinical endpoints. When biomarkers mature into surrogate endpoints, these surrogate endpoints can substitute for clinical endpoints in confirming studies. Surrogate endpoints that are observable prior to the ultimate clinical endpoint are of great value because they can shorten the duration of confirming studies, thereby abbreviating the duration of the development and approval of a drug, thus bringing treatment to patients before the information on clinical outcomes becomes known.

While it is appealing to employ surrogate endpoints to improve clinical trial efficiency, some concerns have been expressed about this approach (4, 5), fueled by notable failures. The most notorious of these was the Cardiac Arrhythmia Suppression Trial (CAST) that demonstrated, though correlated with clinical response, that the suppression of runs of ventricular tachycardia and premature ventricular contractions did not improve but actually worsened survival (6). The approach of using biomarkers as surrogate endpoints functions best when adequate and appropriate safety data are obtained and provisional approval is granted pending follow-up with Phase 4 studies. Despite notable failures of biomarkers used as surrogate endpoints, some such as blood pressure and lipoprotein profile are accepted by clinicians and regulatory agencies as a basis for use and approval.

In addition to being useful in drug development, biomarkers have utility in direct patient care as diagnostic tools, as disease staging tools, as predictors of clinical response to treatment, and for monitoring the progress of treatment.

Few biomarkers will become surrogate endpoints. However, characteristics supporting a biomarker maturing into a surrogate endpoint are (a) biologic plausibility, (b) successful application in prior clinical trials, and (c) presence of significant risk–benefit considerations. Table 17.1 presents a summary of these considerations

as stated by Temple (7). It should be noted that the characteristics that are desirable in a surrogate marker should also be applied when one is deciding which biomarker to employ in a clinical trial or simulation. It will not be possible to have uniform criteria that apply to all biomarkers becoming surrogate endpoints.

17.3.3.2 *Statistical Criteria for Surrogacy*

Statistical criteria and the conceptual framework for surrogacy were initially stated by Prentice (8): "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint." This was later modified by Freedman et al. (9) and further refined by Buyse and Molenberghs (10). We concentrate on the Buyse–Molenberghs approach here. The statistical criteria for

TABLE 17.1   Support for Surrogates

Factor	Favors Surrogate	Does Not Favor Surrogate
Biological plausibility	Epidemiologic evidence extensive and consistent Quantitative epidemiologic relationship Credible animal model shows drug response Well understood disease pathogenesis Drug mechanism of action well understood Surrogate relatively late in biological path	Inconsistent epidemiology No quantitative epidemiologic relationship No animal model Pathogenesis not clear Novel actions not previously studied Surrogate remote from clinical outcome
Success in clinical trials	Effect on surrogate has predicted outcome with other drugs of same pharmacologic class (supports surrogate in class) Effect on surrogate has predicted outcome in several classes (supports more general use)	A negative outcome without clear explanation Inconsistent results across classes
Risk–benefit, public health considerations	Serious or life-threatening illness and no alternative therapy Large safety database Short-term use Difficulty of studying clinical endpoint (rare, delayed)	Nonserious disease and alternative therapy with different pharmacologic action known to affect outcome Little safety data Long-term use Easy to study clinical endpoint (short-term study) Long-delayed, small effect in healthy people

Source: From Ref. 7, used with permission.

surrogacy are indirect evidence that a biomarker may be suitable as a surrogate; therefore, other criteria must also be met, which have been previously stated.

To explain the framework for statistical surrogacy (see Figure 17.1), define Z as the treatment, S as the biomarker, and T as the true clinical endpoint. The effect of the treatment (Z) on the biomarker (S) is called α; the effect of the treatment on the clinical endpoint (T) is called β; and the effect of the biomarker (S) on the clinical endpoint (T) is called γ (10). Statistically speaking, the biomarker can only be used as a surrogate endpoint if an estimated treatment effect on S ($\alpha \neq 0$) can be used to predict a treatment effect on T ($\beta \neq 0$) and if no treatment effect on S ($\alpha = 0$) predicts no treatment effect on T ($\beta = 0$) with sufficient accuracy (10).

From Figure 17.1 consider the situation where S and T have a bivariate normal distribution and the data are obtained from a single study. One can model the relationship between S and T and Z as three distinct linear regressions:

$$S = \mu s + \alpha Z + \varepsilon' s \tag{17.1}$$

$$T = \mu t + \beta Z + \varepsilon' t \tag{17.2}$$

$$T = \mu + \gamma S + \varepsilon' \tag{17.3}$$

where α, β, and γ are slopes; μs, μt, and μ are intercepts; and $\varepsilon' s$, $\varepsilon' t$, and ε' are the residual random effects. A further multiple linear regression relationship can be expressed as:

$$T = \mu' + \beta_s Z + \gamma_z S + \varepsilon' \tag{17.4}$$

where μ' is the intercept, β_s is the slope of the relationship between Z and T in the presence of the concurrently modeled S to T relationship, and γ_z is the slope of the relationship between S and T in the presence of the concurrently modeled Z to T relationship.

A framework for the added complexity of multiple trials can also be incorporated by including between and within study random effects.

One source of evidence for surrogacy is the association between the biomarker and clinical endpoint at the level of the individual. We expect that there will be a high degree of association between the biomarker and the clinical endpoint at
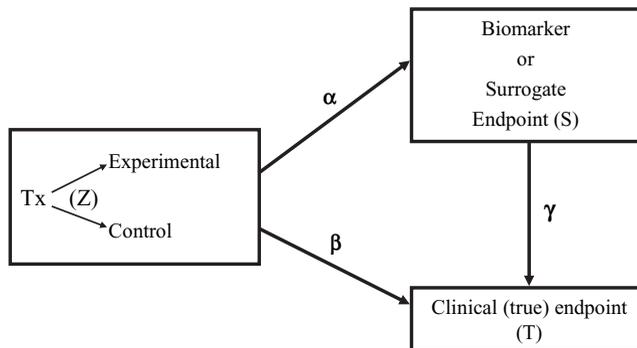
FIGURE 17.1 Representation of the relationship between a biomarker, surrogate endpoint, and clinical endpoint (modified from Reference 10).

the level of the individual if there is some biological pathway from the biomarker to the clinical endpoint. However, it has been demonstrated that a high degree of association does not a surrogate make (4). The other source of association between the biomarker and the clinical endpoint that is needed to establish a biomarker as a surrogate is at the level of the population.

At the level of the individual above, the association between the biomarker and the clinical endpoint could be estimated by γ_z from Eq. (17.4). Another possible variable at the individual level would be the squared correlation between S and T after accounting for the treatment effect. For a biomarker to become a clinical endpoint, we would require that the squared correlation be large, that is, close to 1.

In a seminal paper, Prentice (8) proposed the following criteria for surrogacy:

1. The treatment must have a significant effect on the biomarker; $\alpha \neq 0$.
2. The treatment must have a significant effect on the clinical endpoint; $\beta \neq 0$.
3. The biomarker must have a significant effect on the clinical endpoint; $\gamma \neq 0$.
4. The full effect of the treatment on the clinical endpoint must be captured by the biomarker; $\beta_s = 0$ (see Eq. (17.4)).

Major problems with the Prentice criteria are that they require the biomarker to capture the "full effect" of the treatment on the final clinical outcome in order to become a surrogate (thus, β_s from Eq. (17.4) would have to be 0). Therefore, these criteria are stringent and would only be useful for rejecting a poor marker when the statistical test on β_s, the treatment effect on the true endpoint, resulted in a value that did not equal 0.

To deal with this the problem of the Prentice criteria, others (9, 10, 11) proposed that one estimate the proportion of the treatment effect (PE) captured by the surrogate. Thus, from Eq. (17.2) and (17.4) one can estimate this as

$$PE = (\beta - \beta_s)/\beta \tag{17.5}$$

where PE is the proportion of the effect that can be explained by the biomarker (9). From Eq. (17.2) and (17.4) it can be seen that if the proportion of the effect of the treatment that is explained via the biomarker, γ_z, is large, then β_s will be small and PE will be large. A large PE would be a desirable property for a biomarker that is to be used as a surrogate. A question remains of, "what is large."

Finally, Buyse and Molenberghs (10) have proposed that one estimate the "relative effect" (RE). This is done by dividing β from Eq. (17.2) by α from Eq. (17.1). Of importance here is the precision of the RE. If the precision is high, then the biomarker may be a good surrogate. A major problem with the RE is that it requires large amounts of data to obtain a precise estimate. See Buyse and Molenberghs (10) for detailed application of statistical validation of surrogate endpoint.

17.4   ANALYTICAL INTEGRITY

Analytical integrity is needed for a biomarker to be useable. The biomarker must have a degree of sensitivity, specificity, precision, and reproducibility to accomplish

the job at hand. When evaluating reproducibility one must ask how reproducible the biomarker is from hour to hour, week to week, and year to year in the presence of both health and disease. Within the context of reproducibility one must further answer the question: "Are there circadian rhythms?" Finally, we must consider how much of a change in biomarker would take place between health and disease and between treated and untreated patients. These factors must be taken into consideration when determining to use a biomarker.

17.5 TECHNOLOGIES FOR BIOMARKERS

Biomarkers can be categorized into two major groups—those that are biomolecular/chemical and clinical markers. Clinical markers include such things as blood pressure, neurological scoring scales (Glasgow Coma Score), and nuclear imaging. Over the last decade the array of biomarkers available to the researcher and practitioner has increased to a very large degree. These include electrophysiological, imaging, genomic, proteomic, lipomic, metabolomic, flow cytometry, and molecular diagnostics. It is expected that novel biomarkers will continue to be discovered and applied to drug development. The most common technologies used for biomarkers include standard chemistry, radiology and imaging, genomics and genetics, proteomics, and metabolomics.

Standard chemistry biomarkers have the longest standing history and continue to have the most widespread use in investigational new drugs (INDs) and new drug applications (NDAs). These types of biomarkers have long been used for patient inclusion or stratification, evaluation of clinical response, evaluation of drug toxicity, and biomarker adjusted dosing strategies. Several of these, such as lipoprotein profile and serum glucose, have made the transition to surrogate endpoints for clinical trials when confirming efficacy. Some biomarkers used previously for one purpose are sometimes applied for novel purposes. One such example is cardiac troponin (cTnT), which has been historically used for the diagnosis of myocardial infarction but has recently been used for the estimation of cardiotoxicity of drugs such as doxorubicin. Cardiotoxicity related to doxorubicin administration is insidious. Researchers (12) have demonstrated that increases in cTnT concentrations and cardiac lesion scores increased with increasing exposure to doxorubicin. It was concluded that cTnT released from doxorubicin damaged myocytes, and measurements of serum levels of cTnT appeared to be a sensitive means for assessing the early cardiac effects of doxorubicin (12).

Radiology and imaging biomarkers include radiographs, ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), single proton emission tomography, and echocardiography. The application of this technology has become an area of great interest in recent years, as evidenced by the number of publications and conferences on the topic. Of importance when these methods are applied as biomarkers is image literacy because for some of the methods it is often difficult to find radiologists with the requisite skills to read the image(s). In multicenter clinical trials, there is a multiplicity of sources of data generation and therefore interpretation can be a source of variability. Therefore, for consistency, centralizing the interpretation of the imaging is used to decrease variability in image reading by having only a few specially trained individuals read the images. When there are a series of images

in the same individual, it would be advisable to have one reader for the series of images. The current regulatory guidance in this area states that the reader must be independent of the study and blinded (see *Draft Guidance for Industry, Medical Imaging for Drugs*).

It is expected that the application of imaging as a biomarker technique will continue to see growth in the future. For example, PET has been used to model the progression of Parkinson's disease. K_i is a rate constant that describes the rate of uptake of tracer into neurons. The tracer is preferentially taken up by active neurons in the brain; therefore, K_i is used as a marker for the number of functioning neurons in the brain. K_i was found to be 0.0054/min in Parkinson's patients and 0.0101/min in healthy control subjects (13). MRI has been used to show a greater decrease in brain volume in Alzheimer's disease patients when compared to normal controls (14).

Genomics/genetics biomarkers have the potential to be used to identify patients at risk of disease, predict treatment response, predict adverse events, and predict exposure by identifying biotransformation enzyme activity classification via genetic polymorphisms of drug metabolism (see Chapters 18 and 19). However, the widespread application to improve patient care is still several years away. The most widespread application of genomics/genetics to date has been the typing relative to drug metabolism. Examples of the use of genomic information applied to drug development and patient care are trastuzumab antibody and imatinib mesylate (15).

The history of the development of imatinib provides an informative drug development example of the application of pharmacogenetics to affect a specific genetic target. It starts with the discovery of the between chromosomes 9 and 22 translocation of a bcr-abl fusion gene in the late 1980s. This abnormality is present in 95% of chronic myelogenous leukemia (CML) patients and is known to be a significant contributor to the disease. This gene produces a protein with increased tyrosine kinase activity. Imatinib was developed as a specific tyrosine kinase inhibitor designed to block the ability of bcr-abl to phosphorylate its substrate. This is an example of the development of a compound for a selected gene target (15).

Currently, the most widespread application of genomics/genetics to drug development relates to the polymorphisms of drug metabolism. For example, CYP2D6 is known to metabolize approximately 25% of marketed drugs and has about 70 known mutations. Of these mutations, at least six have been demonstrated to have no enzyme activity while others demonstrate reduced activity (16). There are four phenotypic subcategories of metablizers; poor (PM), intermediate (IM), extensive (EM), and ultrarapid (UM). This is of critical importance because if clearance is much lower in subpopulations where the mutation is present, then systemic exposure to drug would be greater, resulting in increased toxicity in the subgroup. Subgroup-specific dosing strategies would need to be implemented for this subpopulation.

Array profiling is a novel technique that holds much promise but its role in the clinical and drug development setting is yet to be demonstrated (see Chapters 18 and 19).

17.6 MODELING AND BIOMARKERS

The net worth of biomarkers increases significantly when they are integrated in some form with PK, PD, or outcomes models (defined below). These models are

most often structured as a system of differential equations where biomarkers are often the dependent variables in PD models. PK models can be developed and estimated that include covariates such as sex, race, size, and renal or hepatic function. These PK models generate estimates of concentration-related exposure, which are improvements when compared to dose alone. Exposure can be expressed in many forms but commonly includes area under the plasma or serum concentration–time profile, the maximum or minimum concentration during a dosing interval, or the time above some threshold concentration. These concentrations are usually assumed to be correlated with a concentration at a more remote site of action. Exposure variables can then be related through some function to the PD biomarker. The functional exposure–biomarker relationships are reviewed in detail in other chapters of this book (see Chapters 20–26) but often take the form of a linear or log-linear relationship, an E_{max} model, a Hill equation, or an indirect response model. When any of these models are estimated, then the relationship between dose and the PD biomarker can be predicted by linking dose–PK/PD–biomarker.

Biomarkers are of significant value when they are related to some patient clinical endpoint. This relationship is termed an outcomes model. This clinical endpoint can be either a positive clinical response or some adverse advent. An outcomes model translates some surrogate endpoint or biomarker (QTc, blood pressure, international normalized ratio, etc.) into a clinical endpoint such as cure or no cure, improved versus worsened, survival, time to event, disease progression, or wellness score. Although outcomes models are important, they are less often available for use or application than are PK or PD models and therefore their development is one of the greatest areas of need in pharmacometrics. Outcomes models can be time to event models such as Kaplan–Meier curves for which hazard functions can be estimated. The hazard function is a differential equation that, when integrated, links the PD biomarker to the outcome in a time-dependent manner. Discrete outcomes can also be modeled as logistic regression or discriminant function models, where the biomarker at some exact moment in time is related to an outcome. The disease progression model has recently been shown to be useful in relating drug administration to outcomes (17). Other models that should be considered are disease tolerance models that can be applied to such outcomes as tumor or viral load.

Once developed, these integrated models could be used for several purposes. When combined with Monte Carlo simulation, biomarker models can aid in designing clinical trials that are efficient, powerful, informative, and robust. This integration will continue to improve as mechanism-based models of disease are defined, mechanism-based therapeutic interventions are developed and described, and the relationships between drug exposure and clinical response and toxicity are defined. Functional genomics, proteomics, and lipomics will provide support for defining each of these three factors.

When integrating PK, PD, and trial simulations with target biomarkers, one can test previously untested study designs, dose levels, and/or competing dosing strategies. These powerful trials result in a greater probability of demonstrating effectiveness as part of a clinical program within the context of an efficient development program. The linking of the pharmacokinetics, pharmacodynamics, and biomarkers via simulation can also aid in making informed go/no-go decisions.

Integrated biomarker models can provide a mechanistic link between the dose and effect, thus becoming part of the scientific rationale for drug use and approval.

While biomarkers may be employed in conjunction with pharmacokinetics and pharmacodynamics to improve isolated clinical trials, their application to the overall drug development process, especially within the context of the US Food and Drug Administration Modernization Act of 1997 (FDAMA), should bring added efficiencies. Under this provision a fast track to approval has been described, which comprises one appropriate and well controlled clinical investigation plus confirmatory evidence that would comprise scientifically sound data from any investigation that provides substantiation of the safety and effectiveness of the drug. This confirmatory evidence can consist of earlier trials, PK or PD data, or other appropriate studies. This strong supporting evidence can include the impact of treatment on a biomarker or surrogate endpoint. This regulation provides strong incentive to develop novel biomarkers to be used in conjunction with pharmacokinetics, pharmacodynamics, and simulation to substantiate the impact of dosing strategies on a biomarker and support the application.

Integrated biomarker models may provide models for the Bayesian individualization of treatments. Within this context the models provide the Bayesian priors when only sparse data are available on a single patient. From the Bayesian priors and the sparse data, individualized patient parameters can be estimated. This approach leads to individualized dosing while taking into consideration the impact of patient characteristics such as demographics or disease classification.

17.7   ESTIMATION OF BIOMARKER MODELS

Biomarker models that integrate pharmacokinetics, pharmacodynamics, and biomarkers are complex because they are based on sets of differential equations, parts of the models are nonlinear, and there are multiple levels of random effects. Therefore, advanced methods from numerical analysis and applied mathematics are needed to estimate these complex models. When the model is estimated, one seeks a model that is appropriate for its intended use (see Chapter 8).

This ability is available in many software programs. NONMEM (Iconus, Ellicott City, MD) has been widely used to estimate population models arising from both sparse and intensely sampled data. Other programs include WinNonMix (Pharsight Corp., Palo Alto, CA), Kinetica 2000 (Innaphase Corp, Philadelphia, PA), and Pop-Kinetics (SAAM Institute, Seattle, WA). ADAPT II and WinNonlin have focused on PK/PD models and have been combined with Bayesian approaches to estimate population models.

17.8   EXAMPLE OF BIOMARKER ESTIMATION AND APPLICATION

Sodium dichloroacetate (DCA) is a small molecule that has multiple effects on intermediary metabolism. Of primary interest in the current example is the ability of DCA to activate pyruvate dehydrogenase, the rate-limiting enzyme for the conversion of pyruvate to acetyl CoA. The pyruvate concentration is, in turn, replenished by oxidation of lactate, thereby replenishing concentrations of the latter. Such a reduction may decrease the morbidity in head trauma, where local (CSF) elevated lactate is thought to be neurotoxic.

In the current example, cerebral spinal fluid (CSF) lactate concentration was considered to be the biomarker. A model linking the PK exposure to the biomarker and another model linking biomarker to clinical response were estimated and then applied by Monte Carlo simulation to evaluate competing clinical trial designs for a Phase 3 study.

In the PK/PD part of this study, 52 volunteer patients received from 1 to 3 doses of DCA ranging from 45 to 150 mg/kg (18). In total, 1041 DCA concentrations of which 284 were of CSF origin and 1052 lactate concentrations of which 312 were of CSF origin were measured. The CSF lactate concentrations were related to the serum CSF DCA concentrations by an indirect physiologic response model, which is schematically presented in Figure 17.2.

$$\frac{dR}{dt} = K_{\text{in}} - K_{\text{out}} \left(1 + \frac{S_{\text{max}} C_p}{SC_{50} + C_p} \right) R \tag{17.6}$$

where R was the response (the lactate concentration), C_p was the concentration of DCA, S_{max} was the maximum effect by which K_{out} could be increased, and SC_{50} was the concentration at which half of the S_{max} occurred.

For the purpose of the analysis, lactate concentrations were assumed be pseudo-steady-state at the beginning of the first DCA dose, so that K_{in} was set to equal the lactate concentration at time 0 multiplied by the estimated K_{out} ($K_{\text{in}} = K_{\text{out}} \cdot$ lactate). A link between the 24 hour post-trauma CSF lactate and 6 month postinjury Glasgow Coma Score (GCS) was also estimated by application of logistic regression to literature data (20). Thus, a PK-exposure–PD-CSF lactate–outcomes model was constructed with the biomarker, CSF lactate, in the center. It must be recognized that this modeling and evaluation of power and efficiency could not be executed without a biomarker.

A Monte Carlo simulation was executed to evaluate the power and efficiency of competing study strategies. To do this several steps are necessary. First, a template data set was constructed for use in NONMEM. This data set had patient demographics (determined from a typical traumatic brain injury population), doses (either a placebo or the prescribed dose), and so on. From these data a predicted CSF lactate concentration was generated at 24 hours postinitiation of drug. From the CSF lactate the probability of a good response in a patient's GCS post 6 months can be generated, so that a patient would have a probability of a "good response" between 0.00 and 1.00. Therefore, from the logistic regression a patient may have a 0.72 probability of a "good response." However, this is simply the probability of a good response, and in real life, patients who have a 0.72 probability of a
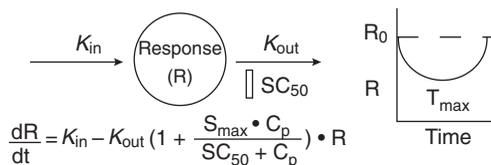
FIGURE 17.2 Schematic representation of the indirect pharmacodynamic response model for dichloroacetate (with permission from Ref. 19).

TABLE 17.2   Assignment of Final Results per Patient in MicroSoft Excel

Patient ID	Placebo or Drug	P of Good Response	Random	Final Result
1	Placebo	0.26	0.77	0
2	Drug	0.76	0.11	1
3	Placebo	0.47	0.63	0
4	Drug	0.86	0.78	1
5	Drug	0.83	0.90	0
6	Placebo	0.39	0.26	1
7	Placebo	0.55	0.32	1
8	Drug	0.31	0.00	1
⋮	⋮	⋮	⋮	⋮
500	Drug	0.78	0.06	1

good response still have a 0.28 probability of not having a good response. To deal with this uncertainty, each patient's probability of a good response along with the patient's treatment group was exported to a MicroSoft Excel file. Here Table 17.2 was created. The P of a good response is the probability of good response from the logistic regression and the logit. Random is a random number with a uniform distribution between 0.00 and 1.00 and was generated from the random number generator in MicroSoft Excel. In Excel, P of a good response is compared to Random by using the conditional if statement. If P of a good response was greater than Random, then Final Result was 1 (indicating a good response) and if P of a good response was less than Random, then Final Result was 0 (indicating not a good response). In the final step, the Drug or Placebo along with the Final Result data were processed in SAS with a chi-square test. For each study strategy, 200 replicates of data were generated, and in the end it was determined how often a given trial structure would be able to detect differences in outcomes between the placebo and treatment groups. So that if in 196 of 200 replicates of a trial, a treatment difference could be demonstrated, then the power of the study would be stated as 0.98.

It was particularly important to determine the impact of several competing dosing strategies and numbers of subjects on power and efficiency. Table 17.3 presents some of the results of these simulations. For the original study design it was intended to enroll 1500 subjects to establish efficacy. The simulation indicated that 500 subjects (250 assigned to placebo and 250 assigned to treatment) would be sufficient to establish efficacy. More patients would need to be enrolled to assess the incidence and severity of toxicity. The use of a biomarker was very important in establishing power and bringing efficiency to a Phase 3 study.

17.9   SUMMARY

Biomarkers have many intended uses. The intended use of the biomarker drives the type and extent of evaluation or validation. If a biomarker is used to select a lead compound for further development, the poor selection of a biomarker has little consequence as far as public health is concerned. The worst case scenario here is the discarding of a good therapeutic candidate. However, when a biomarker is to

TABLE 17.3   Power Estimation of Competing Study Strategies

Study Design Number	Number of Subjects	Dose Levels	Power	Comments
Original	1500	Placebo versus 150 mg/kg	0.99	Expensive with good power
1	500	Placebo versus 150 mg/kg	0.99	Less expensive but toxicity may be unacceptable
2	500	Placebo versus 100 mg/kg	0.99	Less expensive and toxicity expected to be less
3	500	Placebo versus 50 mg/kg	0.98	Less expensive and very little toxicity expected

be employed as a surrogate endpoint in a confirming clinical trial or used to manage direct patient care, there would need to be substantial evidence that the biomarker can predict clinical response.

The development and validation of biomarkers should be part of the overall drug development plan. One key here is "feasible early development." For example, despite all the preclinical pharmacology and promising markers, no surrogate endpoint is validated and therefore any stroke trial requires 5000 subjects. This makes the search for and development of drugs for stroke impractical. Markers should be developed that can be used in both animals and humans. The biomarker should be validated throughout the entire development process. This process of continued validation and strengthening the mechanistic explanation of how the biomarker is involved in the disease–treatment–response may serve to help with a FDAMA (true fast track) type of approach to approval or may be used for a second compound in the pipeline. The judicious use of biomarkers will aid in decreasing the cost of drug development and will improve the quality of direct patient care.

REFERENCES

1. Tufts Center for the Study of Drug Development. Press release November 30, 2001.

2. Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints: preferred definitions and conceptual frame work. *Clin Pharmacol Ther* **69**:89–95 (2001).

3. D. Mildvan, A. Landay, V. De Gruttola, S. G. Machado, and J. Kagan, An approach to validation of markers for use in AIDS clinical trials. *Clin Infect Dis* **24**:764–774 (1997).

4. T. R. Fleming and D. L. DeMets, Surrogate endpoints in clinical trials: Are we being misled? *Ann Intern Med* **125**:605–613 (1996).

5. R. Temple, A regulatory authority's opinion about surrogate endpoints, in *Clinical Measurement in Drug Evaluation*, W. S. Nimmo and G. T. Tucker (Eds.). Wiley, Hoboken, NJ, 1995, pp. 3–22.

6. The Cardiac Arrhythmia Suppression Trial II Investigators, Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* **327**:227–233 (1992).

7. R. Temple, Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* **282**:790–795 (1999).

8. R. L. Prentice, Surrogate endpoints in clinical trials: definitions and operational criteria. *Stat Med* **8**:431–440 (1989).

9. L. S. Freedman, B. I. Graubard, and A. Schatzkin, Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* **11**:167–178 (1992).

10. M. Buyse and G. Molenberghs, Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* **54**:1014–1029 (1998).

11. D. Y. Lin, T. R. Fleming, and V. DeGrutta, Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med* **16**:1515–1527 (1997).

12. E. Herman, S. Lipshultz, N. Rifai, et al., Use of cardiac troponin T levels as an indicator of doxorubicin induced cardiotoxicity. *Cancer Res* **58**:195–197 (1998).

13. P. K. Moorish, J. S. Rakshi, D. Bailey, G. V. Sawle, and D. J. Brooks, Measuring the rate of progression and estimating the preclinical period of Parkinson's disease with [^{18}F]/DOPA PET study. *Adv Neurol Neurosurg Psychiatry* **68**:427–431 (1998).

14. M. J. de Leon, J. Golomb, A. E. George, et al., The radiologic prediction of Alzheimer's disease: the atrophic hippocampal formation. *Am J Neuroradiol* **14**:897–906 (1993).

15. B. J. Druker and N. B. Lydon, Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* **105**:3–7 (2000).

16. G. Wilkinsen, Pharmacokinetics, in *Goodman and Gilman's: The Pharmacologic Basis of Therapeutics*, 10th ed. McGraws-Hill, Highstown, NJ, 2001, pp. 15–17.

17. P. L. S. Chan and N. H. G. Holdford, Drug treatment effects on disease progression. *Ann Rev Pharmacol Toxicol* **41**:625–659 (2001).

18. P. J. Williams, J. R. Lane, C. Turkel, E. Capparelli, Z. Dziewanowska, and A. Fox, Dichloroacetate: population pharmacokinetics with a pharmacodynamic link model. *J Clin Pharmacol* **41**:259–267 (2001).

19. N. L. Dyneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **4**:457–478 (1993).

20. A. A. F. DeSalles, H. A. Kontos, D. P. Becker, et al., Prognostic significance of ventricular CSF lactic acidosis in severe head injury. *J Neurosurg* **65**:615–624 (1986).

Analysis of Gene Expression Data

DANIEL BRAZEAU and MURALI RAMANATHAN

18.1   INTRODUCTION

Many drugs exert effects through changes in gene expression and this chapter focuses on aspects of modeling gene expression as a means to better understand the molecular mechanisms underlying drug response. The processes of receptor activation, signal transduction, transcriptional activation and transcription, RNA processing, transport and degradation, translation to protein, and protein degradation can all potentially contribute to the dynamics of gene expression. Although many of the individual processes are delineated in molecular detail, measurements of individual processes are rarely if ever available in the pharmacometric (PM) setting. The challenge in gene expression modeling is to develop parameter efficient models that effectively describe the relationships between drug concentration and effect despite subsuming multiple underlying variables and processes.

18.1.1   Microarrays for Gene Expression

The availability of arrays for gene expression profiling now allows simultaneous measurements of thousands of RNA species from single samples. Arrays for protein measurements are also becoming increasingly available. Gene arrays are now widely employed in basic biomedical research for mRNA expression profiling and are increasingly being used to explore patterns of gene expression in clinical research. The analysis of array data has been the subject of active multidisciplinary biostatistical, computational, and bioinformatics research primarily because the size and dimensionality of array data sets have substantially altered the scope and complexity of the analyses required in experimental settings. This chapter focuses on the array-derived gene expression data but many of the approaches are broadly applicable to analyzing time courses from other high-throughput methods in biological systems. There are several PM challenges in the modeling of gene expression profiles and the last section of this chapter delineates these unresolved challenges.

18.1.2 Technologies

A wide variety of array technologies are available for measuring expression of large numbers of mRNAs (1). These include (a) oligonucleotide arrays synthesized directly on silicon chips using photochemical technology, (b) oligonucleotide arrays synthesized using ink-jet technology, (c) DNA or oligonucleotide arrays immobilized on glass slides, and (d) DNA arrays on nylon membranes. The technologies differ in the substrates (e.g., silicon, glass, nylon) used, the synthetic methodology employed (e.g., polymerase chain reaction (PCR) with UV crosslinking, synthetic oligonucleotides with chemical crosslinking, photochemical synthesis), the length of probe immobilized (PCR products or oligonucleotides), and the readouts employed (fluorescence, chemiluminescence, or radioactivity). Figure 18.1 shows a schematic of the underlying principles arrays. The human genome has approximately 20,000–30,000 genes, which are estimated to produce approximately 45,000 transcripts due to alternative splicing. Whole-genome expression profiling (2, 3) and large-scale mutant mapping (4, 5) have been possible in yeast for a few years and the technology for assessing the transcript levels of all the expressed genes in the human genome is emerging. There are several proprietary array platforms that are now commercially available (e.g., from Affymetrix, Agilent, Amersham, and many others). Nonetheless, despite their widespread availability and use, gene expression profiling with arrays is employed primarily as a research tool because several technical/methodological challenges with measuring gene expression accurately in clinical relevant settings remain: the most significant of these challenges is the poor concordance between the major platforms (6, 7). The accuracy, precision, and methodological issues related to microarray methodology are being defined in systematic efforts by several groups (8–12). Although microarray results can be confirmed and extended with quantitative PCR techniques, this lack of concordance suggests the need for standardization among microarray providers in their methods of image collection, background assessment and subtraction, validation, normalization, and analysis. However, arrays are a relatively new and emerging technology and ongoing research will result in improvements in performance and robustness that will ultimately permit the use of arrays in clinical settings.

18.1.3 Quantitative Real-Time PCR

Quantitative real-time PCR (QPCR) technology provides a means to not only confirm microarray results but also to extend and apply the results. Using fluorogenic probes (13–19) or intercalating DNA dyes, the PCR amplification process can be monitored in real time, allowing for the estimation of initial mRNA concentration for a given gene target based on the amplification profiles. Databases are now available with the necessary fluorogenic probe and primer sequences for a variety of genes in humans, mice/rats, and other species (20, 21). As with microarrays and other single-gene assays, it is necessary to normalize the data for differences in mRNA quality, concentrations, reverse transcriptase efficiencies, and other experimental errors (22–24). In addition, standard curves can be constructed for each gene using known amounts of the cloned PCR product (25, 26). Since real-time QPCR is relatively simple and inexpensive, it becomes possible to extend microarray studies to include more replication, and other important parameters such as differ-
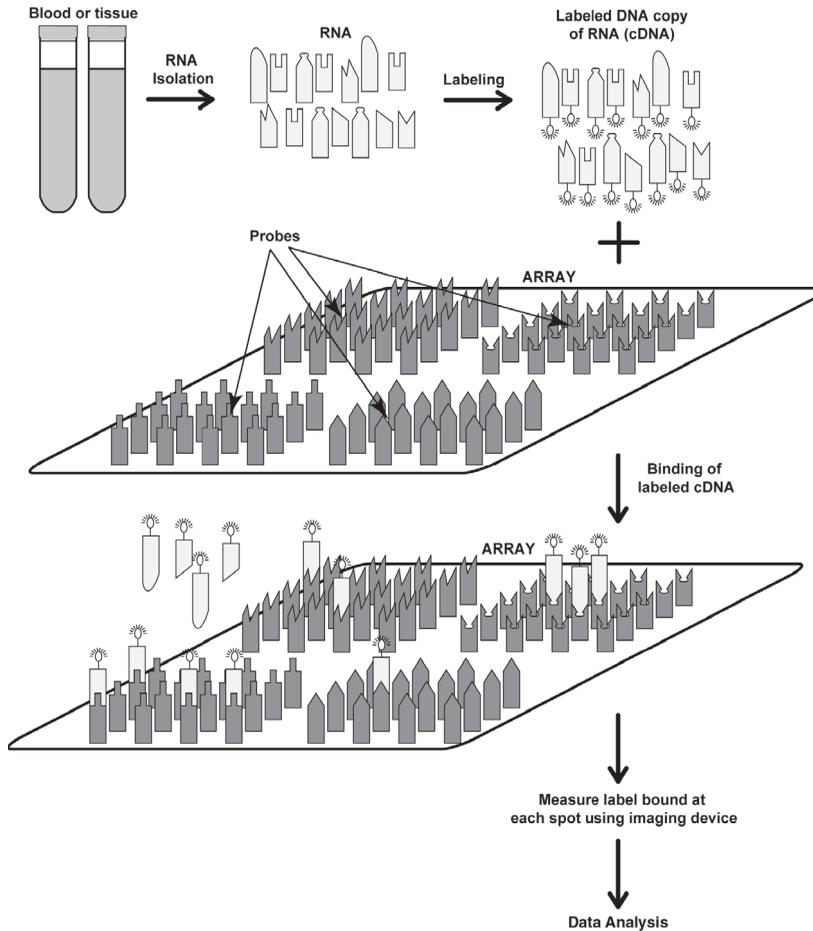
FIGURE 18.1   A schematic of principles underlying gene analysis.

ing dosage regimes and additional time points for specific genes of interest. Real-time QPCR is quickly becoming fundamental to studies ranging from assessments of the kinetics of gene expression in response to specific drugs including interferon-β (27), prednisolone (28), dexamethasone (29), and ciprofloxacin (30) to examining the role of gene expression for transporters (31–34) and/or metabolizing enzymes (35–38) in drug distribution. Real-time PCR is also becoming an important means of genotyping for single nucleotide polymorphisms (SNPs) (39–41).

18.2   MICROARRAY DATA ANALYSIS

18.2.1   Data Normalization

One of the critical first steps in array data processing is normalization. Individual spot intensities are typically normalized on each array to reduce errors associated with experimental methodologies such as starting RNA quality or quantity, differ-

ences in reverse transcription efficiencies and labeling, and hybridization kinetics (42–44). The basic aim of normalization is to correct for experimental bias within each array by normalizing spot intensities to some subset of spots on the array. The simpler normalization approaches range from "global normalization," wherein data from all or nearly all spots are used for normalization, to methods that normalize the data with a small subset of identified "housekeeping genes" or control spots. Often global normalization methods trim the spots to be used in the analysis to eliminate outliers of genes that seem to vary among treatments. Quantile normalization is an increasingly used approach that equalizes the distribution of intensities across all the arrays, ensuring all moments (e.g., including the mean and the standard deviation) are the same after normalization (9, 44, 45). Other recently used techniques include locally weighted polynomial regression (LOWESS), piecewise linearization methods (46), and adaptive algorithms (23, 47, 48).

Ultimately, normalization procedures identify some set of nondifferentially expressed genes across all the arrays of an experiment to serve as a reference set for comparing arrays. With successful normalization, the results of different experiments can be combined for analysis. However, different normalization techniques can have large effects on the outcome of a given experiment and to date there is generally little uniformity among platforms. It is often advisable to evaluate different normalization methods to assess how robust a given outcome is to each method.

18.2.2 Identification of Predictive, Differentially Expressed Genes

The primary advantage of massively multiplexed measurement systems such as arrays is that data on many genes are obtained; however, while some of the genes may be altered by the disease or treatment, many others will be unaffected. Thus, array data contains both informative and uninformative gene measurements and the initial analysis challenge—that of identifying the subset of informative genes—is in principle solvable by selecting genes that meet statistical significance criteria in an appropriate test. The statistical issues are by no means trivial and are complicated by the varied sources of random error and bias and the large number of multiple comparisons involved in array experiments.

In many cases, the initial analysis of array data involves the identification of a subset of genes from the many thousands assayed in an experiment that exhibit significantly altered gene expression due to the experimental conditions. The simplest assessment of significantly altered genes are simple comparisons of fold change using some arbitrarily chosen threshold (generally 1.5–2-fold increase or decrease). The major shortcoming with this simple method aside from a lack of statistical rigor (49) is that ratios mask information concerning the absolute levels of gene expression. More traditional hypothesis-driven analyses include parametric and nonparametric univariate statistical tests (*t*-tests, Wilcoxon) (50–54). Analysis of variance models (55–57) have also been proposed that provide greater generality in model building and experimental design. ANOVA models may even include some preprocessing steps like normalization if the appropriate replication is included in the experimental design. Regardless of the statistical test employed, problems arise when conducting many multiple tests such that as the number of hypotheses tested increases the probability of rejecting a true null hypothesis (Type 1 error) increases. There exist computationally simple corrections (Bonferroni, Sidak) to

deal with the multiple testing problem; however, such corrections are less than ideal for microarray analyses since these methods assume that the multiple tests are independent of one another, unlikely given the many known gene interactions, and the adjustment in significance levels result in a great loss of statistical power. Given that microarray experiments are exploratory in nature, it is generally more acceptable to tolerate some small number of false positives rather than discard some truly significant genes. Permutation-based procedures or bootstrapping techniques have potential application at all levels of microarray analysis and have been used extensively in other fields of genetics and biometry (evolutionary biology, phylogenetics, and population genetics (58)). These techniques allow for the estimation of a null distribution for a variety of test statistics (59–64) in order to assess the significance of individual genes. Permutation-based procedures have been combined with step-down adjustments to control the family-wise error rate for multiple testing (65). These techniques yield significance values with the traditional interpretation: that is, the probability of committing at least one Type I error in the entire data set does not increase with the number of tests. Finally, one permutation-based statistical procedure is specifically available for microarray analysis (statistical analysis of microarrays—SAM (61, 66)). Unlike traditional statistical tests, SAM does not identify a specific list of significant genes; rather, the user can adjust the false discovery rate to assess significant gene lists given different tolerance levels for false positives. Comparisons among permutation-based methods, parametric tests, and traditional nonparametric tests indicate good concordance (64, 67). One of the major limitations of permutation-based methodologies is the granularity of *p*-values that arises when the number of permutations is small due to limited sample size. This is particularly a problem in microarray studies given that the cost of the arrays often limits replication in many studies.

Efficient data storage and retrieval, analysis, statistics, modeling, visualization, and informatics algorithms are the critical tools for biomedical discovery with arrays. A variety of powerful, established tools for biostatistical data analysis (e.g., SPSS, S-Plus, and SAS) and bioinformatics (e.g., GeneSpring, Genomax, and the NCBI web sites' tools) are commercially available. However, the development of visualization, analysis, and modeling tools for time course data for arrays is needed.

18.2.3 Clustering Techniques

The use of clustering techniques is now particularly widespread for examining the temporal dynamics of gene expression, which are frequently necessary to delineate the temporal sequence of transcriptional events that occur in response to a given stimulus. The identification of groups of genes with "similar" temporal patterns of expression is usually a critical step in the analysis of kinetic data because it provides insights into the gene–gene interactions and thereby facilitates the testing and development of mechanistic models for the regulation of the underlying biological processes.

Cluster analysis techniques with a variety of distance measures and decision-generating algorithms have been extensively explored for the analysis of gene array data (68–70). Array experiments in cellular models suggest that certain genes with similar function exhibit similar temporal patterns of coregulation (71–73), although this distinction is not absolute. Generally, the analysis of clusters yields a subset of expected genes with known functions as well as novel or poorly characterized

findings that can provide a basis for further investigation. Each cluster tool and each data filtering/conditioning method within a given technique has the potential for revealing different patterns: this is a frequent source of confusion for some experimentalists.

Clustering algorithms can be classified as either supervised or unsupervised—supervised algorithms require and use more input from the user regarding the underlying structure of the data than unsupervised techniques—and each approach has useful and complementary roles in the analysis of genomic expression profiling data. Supervised approaches (e.g., biostatistical (66), linear or quadratic discriminant analysis (74, 75), neighborhood analysis, and support vector machines (76–78)) are usually more appropriate for analyzing key outcomes and hypothesis testing in well designed, experimental settings. Often the goal of supervised approaches is to identify a subset of the data (genes) that can be used to make predictions or assignments for unknown samples (79). Unsupervised approaches (e.g., hierarchical clustering (73), k-means clustering (80), and self-organizing maps (81)) are more appropriate for data mining and hypothesis generation. We refer the interested reader to the textbook by Webb (82) for an accessible introduction to the mathematical and computation principles underlying these pattern recognition techniques and a comparative assessment.

The goal of clustering is to partition the data set into groups such that members of each group share similarity with each other and are dissimilar with members of other groups. There are two main components to a clustering algorithm: the distance measure and the rules that partition each data point to a group. The partitions produced by a given clustering algorithm are dependent on the distance metric used and must be considered in respect to the distance metric used to generate it (see Ref. 46 for a discussion of different distance metrics). In addition, not all clustering algorithms are deterministic. Membership of any given gene in any particular cluster may be dependent on the initialization parameters. For such algorithms (k-means and self-organizing maps) it is useful to assess the robustness of any given clustering by repeating the algorithm. Hierarchical clustering (HC), k-means, self-organizing maps (SOM), principal components analysis (PCA), and support vector machines (SVM) are some of the commonly used methods for analyzing gene expression data.

18.2.3.1   Unsupervised Approaches

Principal Components Analysis   Principal components analysis (PCA), a technique for reducing the dimensionality of data that is usually performed by singular value decomposition, has also been applied to array data (83–85). The first component from PCA identifies a linear combination of the variables that explains the majority of the variation in the data set and each successive component partly explains the remaining variation in the data set. The components are independent of each other—that is, they are orthogonal. The software program SVDMAN, available for free from Los Alamos National Laboratory (`http://public.lanl. gov/mewall/svdman/`) for DOS and Linux operating systems, provides an implementation of the PCA method for gene expression analysis. In theory, the principal components can also yield composite biomarkers for classification of training sets and class identification because the principal components are an ordered, uncor-

related set of linearly transformed combinations of the original variables, with the first few principal components describing most of the variation in the original data. However, the appropriate component for class identification can only be identified emprically; the linear discriminant function approach, which also uses a linear combination of the original variables, is a more effective method for classification.

Although PCA reduces the number of genes involved, the results largely depend on the data distribution and the variance–covariance of the data. The identified principal components do not always have useful sample prediction capabilities; for example, they often do not capture phenotype structures (86). The poor predictive capabilities of PCA with array data arise because the genes accounting for most of the variance in the data are frequently not the most informative of the class distinction of interest.

Principal components analysis has also been applied to array time series data (83–85) and a limited number of principal components usually accounts for the essential features of the data set, allowing considerably reduced complexity; for example, the sporulation data was modeled using as few as two principal components (83).

By modeling gene expression as Markov processes, Holter et al. (84) extended the principal components analysis/singular value decomposition to estimate the transition matrix for a subset of the principal components. Because the kinetic data is obtained at a limited number of time points, the general problem of computing the transition matrix for an array containing G genes containing G^2 elements and is ill posed. Transition matrices for clustered data and for interpolated time courses have also been examined (87).

Hierarchical Clustering   Hierarchical clustering (HC) is a widely used unsupervised method for microarray data analysis. HC algorithms can be of either agglomerative (bottom-up) or divisive type (top-down) depending on whether they decompose the data set by merging the two nearest individual data points or by splitting a larger group of data points. For gene expression data, agglomerative methods are more widely used. HC algorithms are deterministic though the choice of method (agglomerative or divisive) may generate different patterns. Agglomerative methods for HC produce a series of partitions of the data sets, with the first partition containing individual data points and the last partition subsuming the entire data set.

The HC algorithm requires choice of a distance measure and linkage method. The distance measure quantifies the similarity or dissimilarity between two gene expression profiles. The Euclidean distance and the Pearson correlation (PC) coefficient have been widely used as distance measures to quantify the similarity between profiles. The centered PC similarity measure, r, between any two series of numbers $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ and $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ is the familiar PC coefficient used in linear regression. The distance measure is obtained by subtracting the correlation value from unity. The uncentered PC is obtained from the centered PC by setting the means of $\mathbf{X}$ and $\mathbf{Y}$ to zero. The uncentered PC is defined as

$$r(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \frac{\displaystyle\sum_{i=1}^{N} X_i Y_i}{\sqrt{\displaystyle\sum_{j=1}^{N} X_j^2} \sqrt{\displaystyle\sum_{j=1}^{N} Y_j^2}}$$

The centered PC measure is insensitive to time shifts or translations of the data whereas, the uncentered PC, which is sensitive to time shifts, is usually preferable for most applications. The linkage method determines the basis for assigning the distance or similarity to clusters and data points to accomplish agglomeration; for example, single linkage uses the distance between the two data points in each cluster that are closest and complete linkage uses the two farthest points. Average linkage uses the sum of all the pairwise distances normalized to the number of data points in each cluster. Average group linkage assigns the mean value to each cluster upon merger and the distance between two groups is obtained by obtaining the difference between the mean vectors. Hierarchical clustering generates a hierarchical tree, often referred to as a dendrogram, that highlights the relationships between data points; however, unlike other clustering methods, its does not specify explicit clusters, which the user must be able to obtain by pruning the dendrogram at an appropriate level. The Cluster and TreeView suite, developed by Eisen (73), provides a computational and graphical environment for visualizing clusters in gene expression data and is widely used. The Cluster/TreeView program is available from `http://rana.lbl.gov/EisenSoftware.htm`. The TreeView visualization presents clustered expression data in a combination view that includes both the dendrogram and the heat plot. The dendrogram represents the hierarchy of cluster structures and the heat plot provides information on the expression level changes. The heat plot complements the dendrogram because it is visually very effective and intuitive. Likewise, GeneCluster summarizes results from SOM (73) and J-Express offers visualization clustering results of four major clustering algorithms: hierarchical clustering, SOM, PCA, and k-means (88).

Although HC is widely used for visualizing gene expression data, it has several weaknesses: it is sensitive to noise and to outliers and has a tendency to disrupt large clusters. The HC algorithms use local decisions to identify relationships between data points and dendrogram outputs are obtained even with random inputs. Agglomerative HC produces deterministic results, but "bad" decisions made early on during tree construction cannot be subsequently corrected and as the clusters become larger, the profile of the cluster centroid may sometimes lack any resemblance to the constituent profiles. In addition, the position of patterns within a cluster does not necessarily reflect similar expression profiles. HC is best suited for describing data sets in which the underlying processes exhibit hierarchical relationships and are not ideal when the expression patterns are the result of multiple pathways.

k-Means   The k-means algorithm requires the user to specify the number of clusters to be identified and to estimate/guess the cluster centers. The algorithm randomly chooses points as the centers of the clusters to initiate the process. Each data point is assigned to the closest cluster center and the cluster center is then revised to the center of the points assigned to it. Since each cluster center has moved, membership of each data point within the cluster is then reevaluated. The process is iterated until it converges, that is, no changes in cluster membership and therefore no change in the cluster centerpoints. The problem of identifying the number of centers is difficult and critical since this method is not deterministic. In some circumstances the number of clusters is known a priori (number of time points, tumor classes). A common approach is to use information criteria such as the Schwarz information criterion familiar to those involved in pharmacokinetic/pharmacodynamic (PK/PD)

modeling. Confidence in a given pattern can be assessed by repeating the algorithm a number of times or using a bootstrapping approach using subsets of the data set.

Self-Organizing Maps (SOM)   SOM or Kohonen maps were first proposed in 1995 (89). The SOM algorithm accomplishes the representation of a high-dimensional data set to a low-dimensional (usually one- or two-dimensional) array. Unlike hierarchical and *k*-means clustering, SOM clustering creates plots in which similar patterns occur next to one another—there are "neighborhood relationships" among patterns. This greatly aids in the visualization of the data. SOM requires the user to specify a network of nodes; often this is done by randomly selecting points from the data itself. A data point is randomly selected and the distance between the data point and each of the nodes is calculated. The node closest to the data point is moved in the direction of the data point and the remaining nodes are adjusted depending on their distance from the most proximal node. The *k*-means algorithm arises as a special case of the basic SOM. Like *k*-means, SOM, given the random initialization used to first assign nodes, is nondeterministic.

18.2.3.2   *Strengths and Weaknesses of Clustering Techniques*

Although clustering methods have been widely used in array time series analysis, the majority of these techniques treat time as a categorical or ordinal variable and not as a continuous variable. This distinction is important because the kinetic parameters derived from ordinal variable treatments will not carry meaning except in the case where the time points are evenly spaced.

The majority of clustering techniques currently used for array data analysis are data driven; the initial clustering proceeds with modest levels of user input but effort must then be invested during interpretation. Some of these data-driven methods are also susceptible to noise: many hierarchical clustering approaches lack robustness and uniqueness and can be sensitive to the order of the input and to small perturbations in the data (81). Partition-based approaches can be sensitive to the presence of outliers (90, 91). In addition, the distance measure used can also contribute to the robustness of a given clustering technique. For example, the Pearson correlation, which is widely used as a distance measure for analyzing the kinetics of gene expression because it is insensitive to the absolute magnitudes of the two vectors being compared (68), is capable of identifying visually similar expression patterns but is sensitive to even single outliers. The sensitivity of the Pearson correlation distance measure to single outliers can be reduced by a jackknife procedure wherein each observation is sequentially deleted and the minimum value from the set of correlation values is used for cluster analysis (92). Users should carefully examine the underlying properties of the distance measure when evaluating results.

Clustering techniques, both supervised and unsupervised, have limitations when applied to time series data that are not as problematic or apparent when these same techniques are used for sample-dimension clustering. Typical partition-based methods (e.g., *k*-means and SOM) require the user to provide the number of clusters as a parameter. For time series data, this parameter is difficult to provide because the distinct patterns in a data set may be bridged by a number of intermediate patterns. An example of "bridging" from the Iyer fibroblast response data set (93) is shown in Figure 18.2. In our experience, such bridging is quite common in gene expression profiling data. The intermediate patterns must be forced into
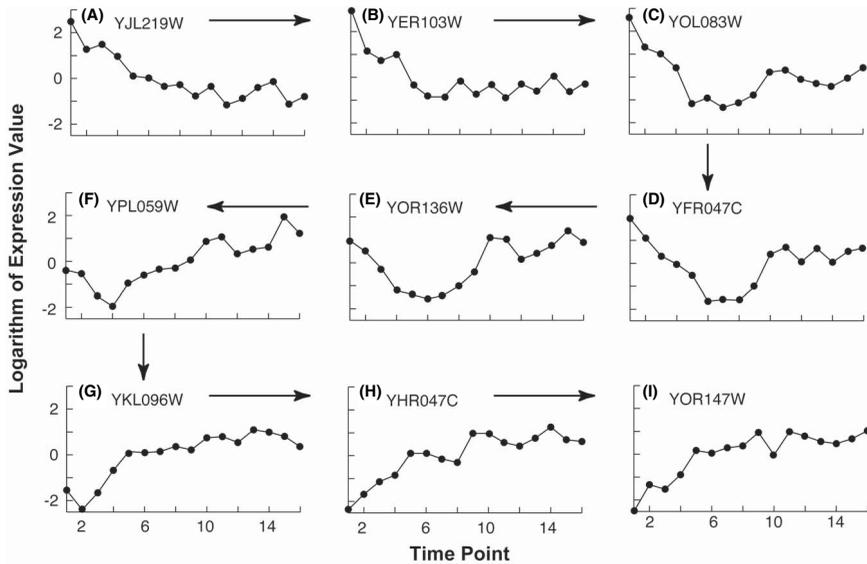
FIGURE 18.2 Example of bridging from the gene expression data set of Iyer et al. (93) on the response of fibroblasts to serum. The qualitatively dissimilar patterns in (A) and (F) can be connected by intermediate patterns present in the same data set.

an existing cluster and generally, existing approaches detect either too few, well populated, noisy clusters or many small but sparse clusters. Hierarchical clustering in the gene dimension eliminates the need to provide the number of clusters as input and creates nested clusters that can be represented as a dendrogram; however, the process of cutting the dendrogram to provide clusters has to be done subjectively by visual inspection. The internal structures of the clusters are difficult to elucidate from the dendrogram.

The CAGED software program (94) is notable among the various clustering approaches because it uses a Bayesian framework for clustering and autoregressive models for representing time series. Generally, because time is treated as an indexed, ordinal variable, the time points have to be evenly spaced for the model parameters to be meaningful as rate constants; it offers a user-friendly interface and excellent biostatistical framework for time series cluster analysis.

18.2.3.3 Evaluating Clusters

Different clustering techniques and distance/similarity measures generally yield different partitions of the data: the number of clusters can differ between methods and a given gene expression profile may be assigned to different clusters. Cluster evaluation metrics can be used for comparing different clustering algorithms and to assess the results from clustering independently of the underlying method. Generally, for temporal gene expression profiles, two complementary types of cluster evaluation metrics are needed. The first type of metric measures the extent to which similar profiles are placed in the same cluster and dissimilar profiles are placed in different clusters, while imposing a penalty for increasing the number of clusters. The second type of metric compares two partitions and measures the similarity

of cluster content or membership, that is, whether the genes assigned to the same cluster by one method share a cluster in the other method.

The Davies–Bouldin validity index (DBI) is an example of a cluster evaluation measure (95). The DBI is the average similarity between each cluster and its most similar one. It is defined as

$$DBI = \frac{1}{N_c} \sum_{j=1}^{N_c} \max_{j \neq k} \left(\frac{(S_c(k) + S_c(j))}{d_{ce}(k, j)} \right),$$

$$\text{where} \quad S_c = \frac{\sum_i \|x_i - c_k\|}{N_k} \quad \text{and} \quad d_{ce} = \|c_k - c_j\| \tag{18.1}$$

S_c and d_{ce} denote the centroid intracluster and intercluster distances, respectively. The intracluster distance for a given cluster is the average of all pairwise distances from points in the cluster to the cluster centroid. The intercluster distance between two clusters is computed as the distance between their centroids. N_k is the number of genes belonging to cluster k, given that a total of N_c clusters are found to exist in the data. A low value of DBI indicates good cluster structure.

The membership matrix **M** for each method forms the underlying basis for comparing for similarity of cluster membership obtained by two different methods, say, P1 and P2. Each term M_{ij} in the membership matrix is an indicator variable that is assigned the value 1 when the gene pair (gene$_i$, gene$_j$) is assigned to the same cluster by both methods and is assigned the value 0 otherwise. The matrix **M** contains all the cluster information and, given this matrix, the clusters and the genes belonging to them can be extracted to generate a 2×2 contingency table:

	0	1
0	n_{00}	n_{01}
1	n_{10}	n_{11}

The total $n_{00} + n_{01} + n_{10} + n_{11} = {}^N C_2$, where N is the total number of genes being partitioned.

The adjusted Rand index (86) measures the extent of agreement between two different cluster structures obtained for the same set of data points. This is a useful measure when comparing two methods producing a different number of clusters. The Rand index is simply the proportion of agreement between the two methods and is defined as

$$Rand\ index = \frac{(n_{00} + n_{11})}{{}^N C_2}$$

N is the total number of genes in the data set. The index ranges from 0 (when the two cluster results are completely different) to 1 (when the two methods agree completely). The expectation of the Rand index for two random partitions is not a constant and it is preferable to use the adjusted Rand index, which corrects the Rand index for the case of random partitions of the data. The adjusted Rand index is given by

$$Adjusted\ Rand\ index =$$
$$\frac{Rand\ index - Expected\ value\ of\ Rand\ index}{Maximum\ value\ of\ Rand\ index - Expected\ value\ of\ Rand\ index}$$

To obtain values for the expected and maximum values, the hypergeometric distribution is assumed. If n_{ij} is the number of genes that are common to clusters i and j from each method, the adjusted Rand index can be calculated from the following formula:

$$Adjusted\ Rand\ index = \frac{\sum_{i,j} {}^{n_{ij}}C_2 - \dfrac{\sum_i {}^{n_i}C_2 \sum_j {}^{n_j}C_2}{{}^nC_2}}{\dfrac{\sum_i {}^{n_i}C_2 + \sum_j {}^{n_j}C_2}{2} - \dfrac{\sum_i {}^{n_i}C_2 \sum_j {}^{n_j}C_2}{{}^nC_2}}$$

The upper bound of the adjusted Rand index is unity and it takes on the value of zero when the similarity between the two clustering methods matches the expectation of the hypergeometric distribution. A higher value of the adjusted Rand index indicates a greater similarity of membership between the clusters of the two methods being compared.

18.2.4 Analyzing Gene Function

The Gene Ontology Database provides the underlying framework of assessing gene function in the results from microarrays. The Gene Ontology Consortium is a collaboration that has developed and maintains structured, controlled vocabularies or ontologies that describe gene products. The vocabularies are species independent and the gene products are described in terms of their associated biological processes, cellular components, and molecular functions. The molecular function describes the biochemical catalytic or binding activity of the gene (e.g., transporter activity is a broad class whereas toll receptor binding is a narrower class); the cellular component describes the anatomical structure that the gene product is associated with (e.g., nucleus, cytoplasm is a broad class, ribosome is a narrower class); the biological processes contain or require ordered multistep assemblies of molecules (e.g., cell growth and maintenance is a broad class, purine metabolism is a narrower class). The gene ontology data sets are freely available from the Gene Ontology web site (`http://www.geneontology.org/`) in flat files, XML, and MySQL formats.

The functions and functional relationships between genes that are statistically significant or comprise clusters can be investigated using the Expression Analysis Systematic Explorer (EASE Version 1.21) software program (96). This customizable software application (available from `http://david.niaid.nih.gov/david/ease.htm`) allows rapid biological interpretation of gene lists and performs theme discovery, annotation, and linking to other online tools such as Database for Annotation, Visualization and Integrated Discovery (DAVID) (97). The one-tailed Fisher exact probability and a variant called the EASE score are statistical measures of

overrepresentation of a class of genes within the total population of the genes in EASE.

Onto-Express is an example of another program that takes lists of genes found to be differentially regulated in array experiments into functional profiles based on gene ontology (98–100). The program provides statistical significance values and graphics of the relative representation of each function class and the gene ontology hierarchical trees. Onto-Express is freely available from `http://vortex.cs.wayne.edu/`.

Although gene ontology mining software tools such as EASE and Onto-Express provide a convenient way to assess biological processes, chemical activities, and cellular localization, it is important to supplement these results by direct visualization of gene expression results on detailed maps of biological processes and biochemical pathways. The Kyoto Encyclopedia of Genes and Genomes (KEGG, `http://www.genome.jp/kegg/kegg1.html`) is an excellent database resource for such analysis (101–104). The KEGG is a suite of databases: the PATHWAY database contains information on molecular interaction networks in biological processes, GENES/SSDB/KO databases contain information on genes and proteins, and COMPOUND/GLYCAN/REACTION databases contain information on chemical compounds and biochemical reactions. The pathway diagram in KEGG represents the interconnections between molecules or, to use an electrical analogy, the wiring network of molecules in biological systems, and can be used in conjunction with microarray and gene expression profiling to assist functional reconstruction. KEGG is fully featured for computing and comparing pathways and contains the necessary binary representations of molecular interactions. Visualization of gene expression data in the context of such diagrams allows users to understand relationships that may initially appear disparate. Software tools for mining the KEGG and gene ontology databases in the context of gene expression profiling data are emerging and examples of such tools include Gene Microarray Pathway Profiler (GenMaPP, which uses its own pathway maps, `http://www.genmapp.org/`) (105, 106) and PathwayAssist (`www.ariadnegenomics.com/`).

18.2.5   Case Studies

18.2.5.1   Case Study 1: Gene Expression Patterns in Interferon-β Treated Multiple Sclerosis Patients

In this case study, we present a detailed step-by-step description of our study of gene expression responses in interferon-β (IFN-β) treated multiple sclerosis (MS) patients (27).

Rationale   The overall aims of this study were to characterize the molecular mechanisms and changes in gene expression patterns associated with IFN-β therapy in MS patients. Recombinant human IFN-β has emerged as the most commonly prescribed form of immunomodulatory treatment for relapsing MS on the basis of several double-blind, placebo-controlled, multicenter trials (107–109). IFN-β reduces relapse rate and slows the progression of disability in relapsing MS. Approximately 30% of MS patients respond well to treatment with IFN-β, whereas the remaining exhibit varying extents of partial responsiveness. Despite this relatively rich understanding of IFN-β signaling, the molecular mechanisms instrumental for its in vivo

therapeutic efficacy in MS are complex and poorly understood. The effects of IFN treatment are complex and the pharmacodynamics of IFN-β at the genomic level in humans is poorly understood. In MS patients, in particular, the benefit associated with IFN-β therapy is difficult to monitor (110, 111), and the cellular, molecular, and immune mechanisms mediating the clinical effects of IFN-β in MS are poorly delineated. Gene expression methodology is particularly appropriate for assessing the treatment effects of IFN-β because it exerts its effects via transcriptional changes mediated through the Jak-Stat pathway (112, 113).

Study Design Considerations In our approach, we utilized a pharmacodynamic (PD) study design. In studies with multiple genomic endpoints, PD designs are particularly important because each half-life of each mRNA (and protein) can differ considerably: a snapshot at a single time point is unlikely to identify whether the gene is being regulated. The PD study design provides insights into gene orchestration because the order in which genes are turned on and off can easily be visualized. We used an open-label PD study design; peripheral blood was obtained from 14 relapsing–remitting MS patients just prior to and at 1, 2, 4, 8, 24, 48, 120, and 168 hours after intramuscular injection of 30 μg IFN-β-1a. Additional samples were obtained at 3, 6, 12, 18, and 24 months, just prior to the weekly dose of IFN-β-1a.

Sample Processing Considerations The separation of different cell populations is an important consideration in genomics experiments because significant changes in cell numbers can occur during treatment, and these changes can confound the gene expression results. In our case, peripheral blood mononuclear cells (PBMCs) were rapidly isolated using gradient separation on cell preparation tubes (Becton Dickinson). Monocytes were depleted from the PBMCs (plastic adhesion) and total RNA was prepared using the TRI reagent method (Molecular Research Center, Inc.) (114).

Choice of Arrays The GeneFilters GF211 DNA arrays (Research Genetics, Inc.) containing named human genes were used (5184 total spots each containing 0.5 ng of approximately 1000 base long, 3′ end-derived PCR fragment). Each filter contained multiple control total genomic DNA positive control spots and housekeeping genes. The manufacturer's recommended protocols were used (`http://www.resgen.com`). This array uses ^{33}P radioactivity for quantification.

   The choice of array is a critical decision because each platform differs considerably in the array formats and labels, methodology, and equipment employed. In our case, we had limited amounts of RNA from patients and our choice of arrays was driven by the high sensitivity and linearity of radioactive readouts and the relatively modest equipment requirements for this platform.

Analysis In the first step of the analysis, the images from arrays were imported directly into the Pathways 4.0 software program obtained from the manufacturer of the array (Research Genetics, Huntsville, AL) and aligned, gridded, and quantified. In this step, the software quantifies the intensities of the spots on the arrays, maps the spot intensities to the genes on the array, and generates a spreadsheet containing intensities. Normalization can also be conducted at this stage and we

opted for global normalization, which uses the intensities across all the spots on the arrays as the normalization factor and also corrects for intensity ranges. The global normalized data were exported as a text file for use in subsequent analyses, which were conducted in other software. Subsequently, we also examined quantile normalization for certain pattern recognition analysis (9, 44, 45).

Our initial analyses focused on biostatistical analysis of key genes that were known from other work to be either modulated by IFN-β or involved in IFN-β signaling (27). Statistical analysis was done in the SAS statistical programs (SAS Inc., Research Triangle Park, NC). Repeated measures analysis with a mixed effect model and linear contrasts was employed for statistical analysis of the time course data for each gene of interest. The PROC MIXED procedure in the SAS statistical program was used. In the statistical analysis, we used $\alpha = 0.01$ rather than $\alpha = 0.05$ to assess significance. We considered using significance analysis of microarrays (SAM) software (66) but because we had a rich time series in a repeated measures design with some missing data, we elected to use the SAS procedure instead. The time profiles for this subset were examined using graphical visualization to assess trends and identify whether heterogeneity of individual responses was apparent. The expression pattern of a subset of these genes was confirmed using real-time PCR.

In the next step of the analysis, we used a variety of pattern recognition tools to investigate the data set. We used self-organizing maps and hierarchical clustering analysis tools using the uncentered Pearson correlation as a distance measure in GeneCluster software (81, 115). Eventually, we used the CAGED approach, which is more suited to time series data clustering (94), to assess the expression patterns in the subset of known IFN-β induced genes. The CAGED analysis indicated that two specific time points, 2 hours and 8 hours, had notable peaks in gene expression. The importance of these time points was also supported by other studies, for example, flow cytometric cell trafficking studies that we had conducted (116). In parallel, we analyzed our data set using novel clustering and visualization algorithms developed in collaboration with our colleagues in computer science (117–119).

We further identified the genes that were significantly changed compared to pre-treatment values at the 2-hour and 8-hour time points using the SAM software (66). The genes that were identified as statistically significant were analyzed for function and gene ontology using EASE Version 1.21 (96), a customizable software application for rapid biological interpretation of gene lists. It performs three basic functions: theme discovery, annotation and linking to online tools such as DAVID (97) and NCBI database. EASE is freely available to nonprofit researchers for use on Windows operating systems at `http://david.niaid.nih.gov/david/ease.htm`. EASE measures the overrepresentation of a class of genes within the total population of the genes in the microarray using the one-tailed Fisher exact probability. We examined the cellular location, biological processes, and molecular activities to determine whether there were specific patterns of overrepresentation.

Analysis is still ongoing. We obtained additional clinical and brain neuroimaging (magenetic resonance images) data in a subset of patients in the course of the study. The current analyses are directed at determining whether there are significant associations between gene expression patterns and clinical and quantitative neuroimaging measures.

18.2.5.2 Case Study 2: Identifying Changes in Vascular Gene Expression Following Nitrate Exposure

The goal was to identify a list of genes with significantly altered expression to understand the molecular mechanisms underlying the multifactorial process of nitrate tolerance in a rat model (120). Global gene expression changes in rat aorta following exposure to nitric oxide donors was examined using cDNA arrays with 5147 rat genes and 384 control DNA spots (GF300 GeneFilters, Research Genetics, Huntsville, AL). Total RNA was isolated from aortas from rats infused continuously with either nitroglycerin (NO donor) or distilled water ($n = 4$ each). Labeling was done with ^{33}dCTP and hybridizations were detected using phosphorimaging. Normalization was done using the simplest method; that is, raw intensity values for each array were normalized to average global intensity values.

The microarray data obtained were quite variable with CVs ranging from 1% to 200% for the 5531 genes. As is often the case, less than 0.5% of the gene signals had CV greater than 100%. Because the goal of these experiments was to identify those genes that exhibit significant differences in gene expression, we employed straightforward t-tests with unequal variance to assess significance and then used permutation-based procedures to assess the levels of false positives given various levels of stringency. The initial t-test identified 447 genes that were altered in their expression by the nitrate treatment at a P level of 0.05. Given this P value, one might expect 276 significant outcomes by chance alone (Type I errors) when conducting a statistical test 5531 times. For the permutations we constructed 35 unique permutations to build null distributions for each gene. Using the simple criteria of accepting as significant any gene (a) whose t-statistic was significant at $P < 0.05$ and (b) whose t-statistic was at the very tail of the constructed null distribution (largest t-value among the 35 permutated t-values), we reduced the list of significant genes to 290. Thus, for the 447 genes originally identified as significant, 157 ($447 - 290 = 157$) yielded higher t-values using randomized data. These genes were removed from the list of significant genes as likely false positives. Finally, the methods of Westfall and Young (65) involving permutation-based procedures combined with Bonferroni-like step-down adjustments were applied to the data set. This technique yields significance values with the traditional interpretation: that is, the probability of committing at least one Type I error in the entire data set does not increase with the number of tests. The Westfall–Young procedure is conservative and when applied to our data the list of significant genes was reduced to 80 genes. This strong control of the family-wise Type I error rate, however, results in increased Type II error. Among the 367 genes ($447 - 80 = 367$) now removed from the list of significant genes, there is a high likelihood of some being significant. Permutation-based methods allow the experimenter to adjust the list of "significant" genes, always with an estimate of the false discovery rate. The level of tolerance for false positives, or conversely the loss of true significant genes, is left to the researcher.

18.2.6 Model-Based Approaches

Model-based or *declarative* approaches, which start with a group of models and use a learning procedure to select one that best describes the data, differ substantively from data-driven or *procedural* approaches, which extract conclusions from observed data using a sequence of steps (121). In the model-based approach, the

structural constraints imposed suppress the effect of noise and allow the results to be easily interpreted because the parameters, which can be viewed as reduced dimensionality representation of the data, reduce complexity of the analysis and have clearly associated meaning.

Appropriate model-based approaches can potentially provide mechanistic insights into the distinct patterns as well as bridging patterns because the response of the models can vary continuously between distinct patterns upon change of parameters.

One of the goals of modeling is to infer the genetic networks and gene–gene interactions from expression data. The major approach has been with Boolean networks (122–127). Friedman and co-workers have used Bayesian networks, which are graph-based models of joint multivariate probability distributions that assess conditional independence between variables, to obtain simpler submodels to describe gene interactions from array data (121, 128). The Bayesian network models have been applied to expression profiles in yeast; several regulatory modules identified were checked against literature reports and testable hypotheses were confirmed empirically using knockout strains (129).

Periodicity approaches have been used for analyzing genes regulated during cell cycles in experiments with synchronized cultures (71, 72). However, these approaches are not generally applicable to nonperiodic data.

Chen et al. (130) proposed the theoretical basis for a series of differential equation-based models for gene expression in the context of array analysis but, surprisingly, did not explore specific applications of the differential equation-based approach to array data sets.

The S-system (or synergistic and saturable system) formalism (131) is a differential equation based approach that has also been applied to genetic, biochemical, and immune network data (132, 133). These systems are nonlinear and both genetic algorithms (134) and linear programming (123) have been used for their analysis. The currently available approaches are not easily applied to large systems and even upon simplification do not yield unique parameter estimates (123).

Despite the availability of a variety of alternative paradigms, compartmental modeling is the dominant approach for understanding the relationships between drug concentration and effect in the pharmaceutical and pharmacological sciences. This paradigm offers a rich body of behaviors for modeling the effects of drugs and models, and model equations can be systematically built from intuitive compartmental elements. A good pharmacokinetic/pharmacodynamic (PK/PD) model could deliver several advantages, including but not limited to (a) a parsimonious, reduced dimensionality representation of the original data, (b) insights into the potential mechanisms or means by which input stimuli are transformed into the time courses of output, and (c) parameters that convey quantitative information that can be used in simulations and what-if analysis or combined with other methodologies already in use for array data analysis (e.g., the parameters of a model could provide a basis for clustering). In the following, we present the results for several simple compartmental models for gene expression dynamics and describe their output characteristics.

The distinction between direct and indirect drug effects is a particularly useful one when modeling gene expression profiles with compartmental models. Direct effects are mediated directly by the presence of the drug in the effect compartment
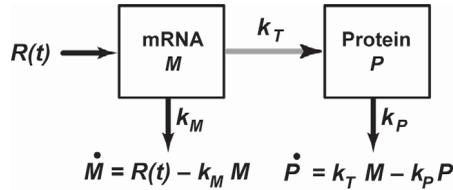
FIGURE 18.3   The Hargrove–Schmidt model and its differential equations. The gray line indicates that information, not mass, is transferred from the mRNA to the protein compartment. The dots in the equations denote time first derivatives.

and because an occupied drug–receptor complex elicits the response, the removal of drug from the effect compartment abrogates the effect promptly. Indirect effect models explain pharmacological effects that manifest when drug concentration in the effect compartment has decayed to negligible levels and with substantial time delays (135, 136). A given drug (e.g., IFN-β) may induce certain mRNAs (e.g., the rapidly induced antiviral genes) via direct effects and others may be better described by indirect effect models.

18.2.6.1 *Models for Gene Expression Dynamics*

The simplest model for gene dynamics is the Hargrove–Schmidt model (Figure 18.2). The Hargrove–Schmidt model is a two-compartment model that assumes information flow from mRNA (M) to protein (P) via a first-order translation rate constant, k_T, and independent, first-order degradation of mRNA and protein with rate constants k_M and k_P, respectively. The system is described by the differential equations and schematic in Figure 18.3.

The tanks-in-series model (Figure 18.4A) assumes a series of N well-stirred compartments with identical residence times (137, 138). The equations describing the tanks-in-series model are

$$\frac{dM_1}{dt} = \frac{DR}{\tau} - \frac{M_1}{\tau}$$

$$\frac{dM_2}{dt} = \frac{M_1}{\tau} - \frac{M_2}{\tau}$$

$$\vdots$$

$$\frac{dM_N}{dt} = \frac{M_{N-1}}{\tau} - \frac{M_N}{\tau}$$

where DR is the concentration of the drug–receptor complex, M_i is the concentration of the signaling species in the ith compartment, and τ is the time constant for each compartment. Mathematically, the bolus or impulse response of the tanks-in-series model is an Erlang distribution, a special case of the gamma distribution with shape parameters restricted to positive integer values; it is also referred to as the stochastic model because the Erlang distribution represents the time required to carry out a sequence of N tasks whose durations are identical, exponential probability distributions. For a unit bolus input dose of receptor occupancy, the level of signaling species in the Nth compartment, M_N, is described by the Erlang distribution equation:

(A)
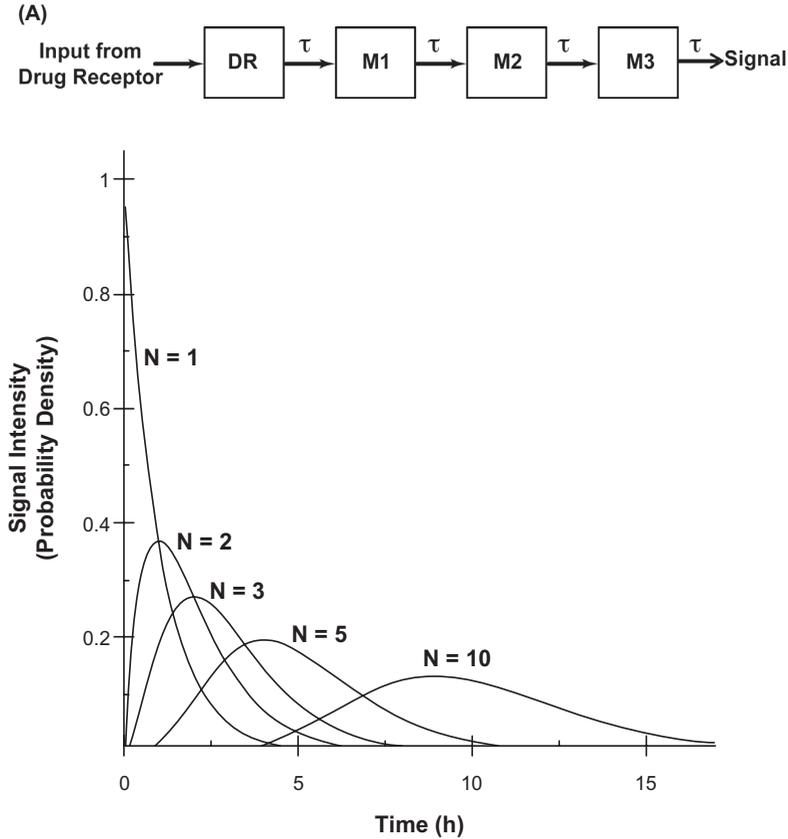
FIGURE 18.4   (A) Representations of the tanks-in-series or gamma distribution model. There are $N = 3$ compartments, each with a time constant τ. The plot shows the bolus response for various N. In the model, Drug, R, and DR are free drug, free receptor, and drug–receptor complex concentrations, respectively; k_1 and k_{-1} are the rate constants for binding and DR dissociation, respectively. M1, M2, and M3 are concentrations of signaling intermediates and h is the Hill exponent. (B) Comparison of the transit compartment model (dashed line) to the tanks-in-series model (solid line). Both models used $N = 3$ compartments and an impulse or bolus input into the M1 compartment; the Hill exponent $h = 2$. The initial condition was M1 = 10 at $t = 0$; the initial conditions in the remaining compartments were zero. Inset is similar in all respects to (B) except that the initial condition was M1 = 2.

$$M_N = \left(\frac{t}{\tau}\right)^{N-1} \frac{e^{-t/\tau}}{(N-1)!}$$

Figure 18.4A shows the output from a tanks-in-series model for a unit impulse or bolus input into the M1 compartment. Each compartment was assumed to have $\tau = 1$. The mean residence time, $\bar{t}$, in an N-compartment tanks-in-series model is simply the sum of the residence times in each compartment and $\bar{t} = N\tau$. The change in the variance of the signal upon passage through the compartments is given by $N\tau^2$. In residence time analysis, it is conventional to normalize the time values through a simple division by the mean residence time value. The mean and variance of the normalized residence times of a tanks-in-series model are $\mu = 1$ and $\Delta\sigma^2 = 1/N$,
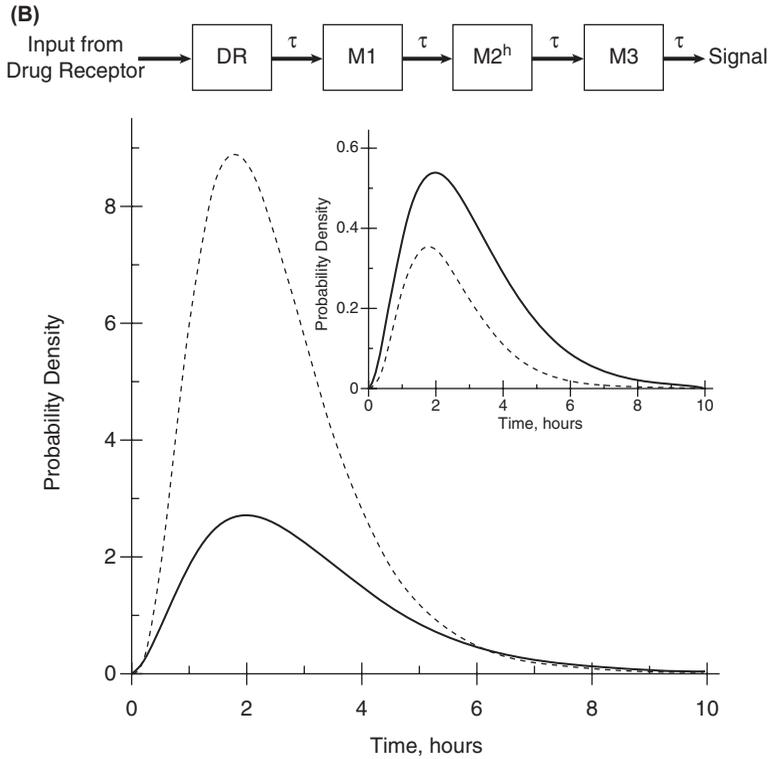
FIGURE 18.4 *Continued*

respectively. Thus, the variance of the signaling times in the tanks-in-series model increases with the number of the compartments, but the coefficient of variation (CV) of the signaling times decreases.

The transit compartment model (Figure 18.4B) is an extension of the tanks-in-series model in which one (or more) of the signaling compartments incorporates nonlinearity via a Hill exponent h (138). The equations are given by

$$\frac{dM_1}{dt} = \frac{DR}{\tau} - \frac{M_1}{\tau}$$

$$\frac{dM_2}{dt} = \frac{M_1}{\tau} - \frac{M_2}{\tau}$$

$$\frac{dM_3}{dt} = \frac{M_2^h}{\tau} - \frac{M_3}{\tau}$$

where DR is the concentration of the drug–receptor complex, M_i is the concentration of the signaling species in the ith compartment, τ is the time constant for each compartment, and h is the Hill exponent. Figure 18.4B compares the transit compartment (dashed lines) to the tanks-in-series model (solid line) and highlight the effect of the Hill exponent-derived nonlinearity. The simulations in Figure 18.4A and Figure 18.4B differ only in the initial conditions used: the initial values of M1

were 10 units and 2 units, respectively. The nonlinear term with $h > 1$ has the effect of amplifying large values and attenuating low values and provides considerable flexibility in fitting certain data sets.

We have developed a model called the dispersion model (Figure 18.5) and compared it to the tanks-in-series and transit compartment models (139). The parameters of the dispersion model estimate the relative roles of diffusion, convection, and chemical reaction in signal transduction. We found that the dispersion model was capable of simultaneously fitting mRNA and protein dynamics for tyrosine aminotransferase (TAT) after methylprednisolone treatment from a published PD study quite well (140).
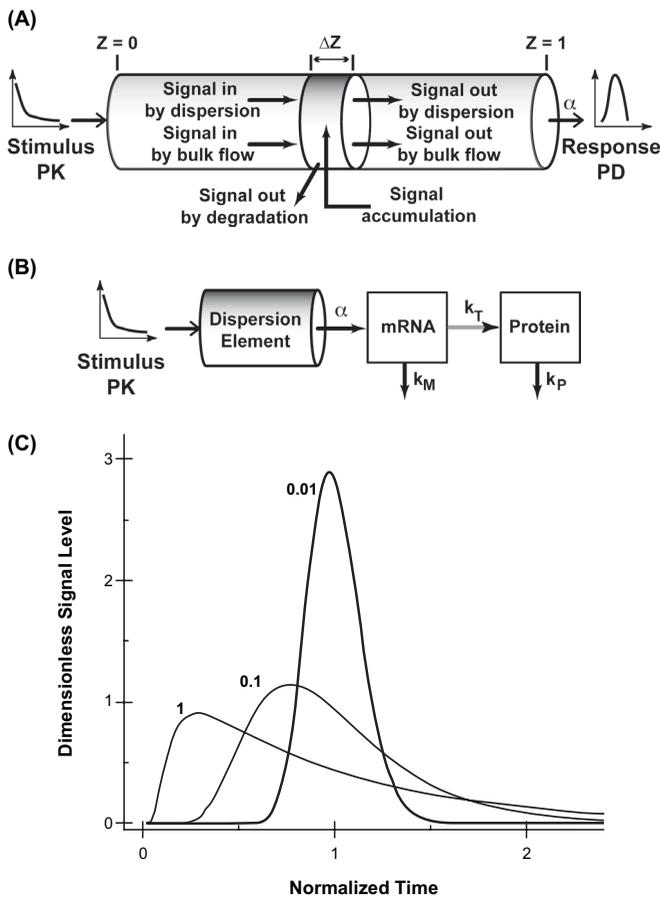
FIGURE 18.5 The mechanistic basis of the dispersion element. The three dispersion element parameters are D_N, a nondimensional dispersion number that measures the rate of signal diffusion relative to convection; τ, the apparent mean residence time; and α, the "signal-to-transcript" conversion parameter. For the mRNA and protein compartments in the Hargrove–Schmidt model element, k_T, k_M, and k_P are rate constants for translation, mRNA degradation, and protein degradation, respectively. The gray line indicates that information rather than mass is transferred from the TAT mRNA to the TAT protein compartment (140).

In the analysis of residence time distributions, it is conventional to normalize time using the mean residence time of a noneliminated bolus input. This makes the normalized mean residence time (μ) of the signal in the dispersion model $\mu = 1$.

For a dispersion model with closed boundary conditions, the change in the normalized variance, $\Delta\sigma^2$, where σ^2_{out} and σ^2_{in} are the output and input variances of the signal, upon passage through the transduction cascade is given by

$$\Delta\sigma^2 = \sigma^2_{out} - \sigma^2_{in} = 2D_N - 2D_N^2\left(1 - e^{-1/D_N}\right)$$

These expressions demonstrate that the normalized mean residence time and variance of the normalized residence time distribution increase with increased values of the axial dispersion number D_N. In the limit of $D_N = 0$, the signal is convected and behavior corresponding to the parallel tube model is approximated: the normalized residence time $\mu = 1$ and $\Delta\sigma^2 = 0$. For very large values of D_N, the behavior corresponds to a single well-mixed compartment.

18.2.6.2 *Opportunities and Challenges in Modeling Gene Expression Data*
Pharmacometrics and PK/PD modeling can provide important insights into gene expression and proteomics data because these approaches can reduce the dimensionality of the problem: time courses of expression are expressed in terms of a limited number of model parameters that are readily interpreted by the users.

As an example of the insights available from the use of these models, consider the observation that the apparent correlation between mRNA levels and protein levels is weak (141–145); typically, only about one-third of the mRNAs found significantly altered in statistical analyses are significantly positively correlated with the levels of cognate protein. A naive analysis of this lack of correlation has been employed to criticize the use of gene expression data, but the underlying mechanisms can be analyzed profitably using PK/PD models such as the Hargrove–Schmidt model. The primary determinants of mRNA levels in the Hargrove–Schmidt model are the transcription rate and the half-life of mRNA, whereas the protein level depends on the mRNA levels, the translational rate constant, and the half-life of protein. According to the Hargrove–Schmidt model, mRNA level will be proportional to the level of the corresponding protein at steady state and a strong correlation could be expected if steady-state levels of an mRNA and its cognate protein are examined (146–148). For a given gene product, the strength of the correlation between the mRNA and its cognate protein will depend on the delay between mRNA and protein profiles: if there is a lag between the mRNA and protein compartments, the correlation across time points will be poor because the mRNA could have decayed while protein levels are just increasing. Thus, for a family of closely related, constitutively expressed proteins, the combined effects of mRNA and protein half-lives and translational constant may be sufficiently close to provide strong correlations for different mRNA–protein pairs. However, because the transcription rates and the half-lives of individual mRNAs and proteins vary considerably, this does not necessarily imply a strong correlation in expression of a random selection of mRNA–protein pairs at steady state. Differences in mRNA and protein half-lives are also likely to cause poor correlations if mRNA–protein levels are monitored under transient conditions that deviate significantly from steady state. If the rate of protein production from translation is relatively small compared to the total size of

the protein pool, the correlation between protein and mRNA is likely to be poor because large changes in mRNA levels may cause only small changes in protein levels. Likewise, if only a small fraction of the mRNA pool effectively contributes to protein levels, the correlation is likely to be poor as well. Thus, modeling highlights the many quantitative arguments for both the existence and the absence of correlations between mRNA levels and protein levels. Furthermore, because the mRNA and protein compartments are causally linked, the relationship between mRNA and protein can be determined only in a PD study design; cross-sectional studies at a single time point are insufficient.

However, the PK/PD modeling community is still struggling to handle gene expression data effectively, largely because compartmental approaches require very high levels of supervision and the system identification can be very time intensive even for data sets containing only a limited number of PD endpoints. Elegant system analysis software tools such as ADAPT, SAAM II, and WinNonlin (to name a few) make compartmental modeling accessible for many users but also require high levels of user intervention, and the fitting of thousands of mRNA expression profiles to even a limited set of model types, while technically feasible, is inconvenient enough to be impractical. Typically, the model selection step is the most difficult to automate.

However, the advantages of (valid) mechanistic models are many: models are more meaningful and easier to interpret, more likely to be robust at predicting responses beyond those studied, better suited for simulations, design of studies (using tools such as optimal sampling theory), and applications such as adaptive feedback control strategies (e.g., "clamping" studies, "concentration- and/or effect-controlled" trials, or algorithms for individually optimizing therapy). For these reasons, new strategies and software packages for modeling gene expression profiles are urgently needed and present critical challenges for PK/PD modeling research. In the following, we present some key challenges and design considerations for the development of compartmental modeling techniques for pharmacogenomics.

The overall strategy for gene expression data requires a robust implementation of user-specified, *generalized* models that can be expected to work *autonomously* with sparsely sampled data sets. The robust generalized approach used should eliminate, if possible, the need to formulate compartments and their connectivity models for each gene. Instead, the user should be able to identify functions that, preferably, are simple yet general and parameter efficient, to describe the growth and decay processes. In contrast, most specific PK/PD and indirect effect models (135, 136) require gene-individualized models, parameter specifications, and assessments of fit. Specific PK/PD and indirect effect models are potentially very useful for limited subsets of interesting genes and for cluster means, and the implementation should be structured so that users will be able to extract the formatted inputs needed and pass the inputs to full-featured programs (e.g., ADAPT (149)) that implement specific PK/PD models.

As an example of a generalized model, consider the schematic in Figure 18.6 for a canonical model that describes drug effects on gene dynamics. The $K_{signal}(t)$ and $K_{decay}(t)$ are referred to as the delay functions for the signal and decay processes, and for the gene expression profiles, the convolution integrals describe how the initial stimulus $S(t)$ evolves during the transduction process to cause changes in the rate of transcription and the rate constant decay of mRNA levels.
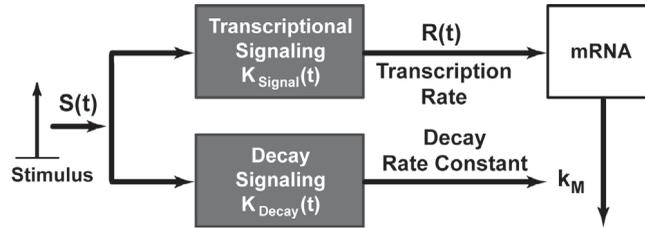
FIGURE 18.6   Schematic of the growth and diffusion with distributed time delay approach. For a bolus, $R(t) = K_{signal}(t)$ and $k_M = K_{decay}(t)$.

One concern with the generalized approach is that if very complicated functional forms are selected, the models may become overparameterized and it may not be possible to discriminate between different model types. The other concern with the generalized approach is that it might be insufficient for characterizing the complex repertoire of gene expression responses. The problem is a challenging one and it may not be a simple task to find relevant functions. The alternatives to the generalized approach would be to employ SOM clustering with a scale-insensitive distance measure (e.g., Pearson correlation) to identify clusters, the centroid of which can then be fit using a variety of indirect effect or direct effect models. Once model identification is completed on each of the centroids, the model-specific parameters for each gene in the cluster can be obtained using the ADAPT system identification software. The alternative approach would attempt to automate the model identification process: the data for each gene would be iteratively fit to each member of a large library of PD models (ranging from direct effect models, delayed effect, and indirect effect models) using ADAPT or other PK/PD modeling software and use scripts that output the model of choice (using, say, the AIC) for each gene to a database; selected output from each of the other models will also be databased so that the user can manually evalaute them if need. This is a brute force approach and is less preferable because it needs higher levels of computing and storage infrastructure.

A critical component of comodeling multiple outputs is the appropriate weighting of individual observations. The weights must be appropriate for small and large responses *within* an output and the relative weights must be appropriate *between* outputs. Failure of the former standard can lead to regions of systematic error in the fitted function and failure in the latter standard can cause some of the outputs to inappropriately dominate the determination of fitted parameters. However, error variance model selection, as for structural model development, should be guided by parsimony: stay as simple as possible.

In any PK/PD modeling effort, critical quantitative assessments of model misspecification and model goodness of fit are necessary. PK/PD modeling strategies for genomic data sets require an additional layer in the performance evaluation process: "global measures" that evaluate misspecification and goodness of fit across all the genes will be needed in addition to the local model misspecification and fit measures that will be used for model selection and parameter estimation for individual genes.

The Akaike information criterion (or other measures such as the Schwarz information criterion) can be used to assess models for individual genes and also to assess the performance of a given modeling strategy on a given data set. In addition to the mean coefficient of determination (r^2) for each fitted function, and the mean percent relative standard error of parameter estimates, the mean value of the normalized root mean square error and the mean value of the mean error can be used as secondary performance criteria. The root mean square error and the mean error between the data and the selected model will be evaluated at the sampling points for each gene and normalized using the standard deviation for each gene. The CPU time required for fitting each canonical model to the data set will also be estimated and used as a tertiary performance criterion.

Model validation is also necessary in addition to assessments of goodness of fit. The approaches that can be used are similar to those used in population PK/PD analysis. Ideally, a series of informative experiments (learning set) are used to develop and to parameterize the model, and those results are used to predict the responses of subsequent experiments (validation set), which test bias and precision of predictions: at similar experimental conditions (of time, inputs, and/or sequence, as were studied in the learning set), at conditions not previously studied but that are within the range of those in the learning set (interpolation), and responses for conditions that are beyond the range of conditions previously studied (extrapolation). This approach to validation is not always feasible. Other variants include randomly splitting the experimental data into learning and validation sets (applicable if the experimental data set is large enough) or performing "internal validation" (randomly exclude "small" portions of data that are predicted based on a model based on the remainder; replace the excluded data and repeat "many" times).

The problem of providing a multitude of initial values and initial parameter estimates needed for any compartmental model in an automated or semiautomated fashion also needs to be addressed. In principle, one strategy for generating initial values is to normalize mRNA levels to pretreatment values, in which case the initial values can all be set to unity. However, genes with low pretreatment values are problematic and will need to be "tagged" and modeled without normalization to avoid floating point problems; the user should have the option of setting the initial values for these genes either to zero, to the observed pretreatment value, or to a specified lower limit of detection. Alternatively, the user will be able to set all the initial values to the measured pretreatment levels. Although software packages, such as ADAPT, allow estimation of initial conditions, initial parameters will still need to be generated. For automated generation of initial parameter estimates, the asymptotic properties of the functions can potenitally be exploited to obtain initial estimates from the data values at the earliest and most distal time points.

18.2.7   Potential Application Areas

Many of the early applications of gene arrays were in the characterization of the heterogeneity of malignancies. For example, Golub (150) compared the gene expression profiles of acute lymphoblastic leukemia to acute myeloid leukemia; diffuse large B cell leukemias were expression profiled by Alizadeh et al. (151), and hereditary breast cancers with BRCA1 and BRCA2 mutations were investigated by Hedenfalk (152).

Subsequently, gene expression with arrays has been used to characterize the responses to a variety of immunomodulatory therapies in humans. The primary advantages of mRNA measurements is that mRNA changes can precede protein changes and potentially allow earlier assessment of responses to drug therapy. For example, Bohen et al. (153) measured the pretreatment gene expression profiles in follicular lymphomas of patients receiving rituximab, a monoclonal antibody directed against the CD20 protein found on B cells. These authors suggested that the gene expression patterns of lymphomas that were nonresponsive to rituximab therapy were more similar to those of control lymphoid tissues than lymphomas that responded to therapy. Chang et al. (154) used gene expression profiling to biopsy samples from breast tumors prior to treatment with docetaxel, a taxane antimicrotubule agent; they identified a 92-gene panel capable of predicting response to drug treatment with an accuracy of 88% and a sensitivity of 85%. Our group has used DNA arrays to characterize the heterogeneity of treatment responses to interferon-β, the most widely prescribed immunomodulatory drug for multiple sclerosis (27). However, only about 30–40% of patients respond well to interferon-β, and the remaining patients exhibit varying degrees of partial responses. Interferon-β exerts its biological effects by modulating gene transcription via the Jak-Stat and other signaling pathways (113). We characterized the responses to intramuscular interferon-β administration in multiple sclerosis patients using a pharmacodynamic study design over an entire dosing cycle (rather than single point measurements). We found that interferon-β induces gene expression changes rapidly and that wide variations in the pharmacodynamics of expression are present (27).

18.3 SUMMARY

There has been rapid development of technologies for large-scale gene expression profiling at the messenger RNA and protein levels from single samples. At this stage of data analysis, it is critical that investigators be cognizant of the many assumptions that underlie every step of the process from image acquisition through data normalization and data analysis. It is also essential that this information be made available to all so that studies can be compared and ultimately pooled to build wider understanding of the molecular processes.

Although there are currently many analytical and modeling challenges with these data sets, further improvements in these technologies will engender applications in drug development, clinical research, and patient care.

ACKNOWLEDGMENT

REFERENCES

1. R. A. Young, Biomedical discovery with DNA arrays. *Cell* **102**(1):9–15 (2000).
2. G. F. Bammert and J. M. Fostel, Genome-wide expression patterns in *Saccharomyces cerevisiae*: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol. *Antimicrob Agents Chemother* **44**:1255–1265 (2000).

3. C. J. Roberts, B. Nelson, M. J. Marton, R. Stoughton, M. R. Meyer, H. A. Bennett, Y. D. He, H. Dai, W. L. Walker, T. R. Hughes, M. Tyers, C. Boone, and S. H. Friend, Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**:873–880 (2000).

4. P. Ross-Macdonald, P. S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K. H. Cheung, A. Sheehan, D. Symoniatis, L. Umansky, M. Heidtman, F. K. Nelson, H. Iwasaki, K. Hager, M. Gerstein, P. Miller, G. S. Roeder, and M. Synder, Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**:413–418 (1999).

5. D. D. Shoemaker, D. A. Lashkari, D. Morris, M. Mittmann, and R. W. Davis, Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat Genet* **14**:450–456 (1996).

6. P. K. Tan, T. J. Downey, E. L. Spitznagel Jr., P. Xu, D. Fu, D. S. Dimitrov, R. A. Lempicki, B. M. Raaka, and M. C. Cam, Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**:5676–5684 (2003).

7. R. Kothapalli, S. J. Yoder, S. Mane, and T. P. Loughran Jr., Microarray results: how accurate are they? *BMC Bioinformatics* **3**:22 (2003).

8. T. Bammler, R. P. Beyer, S. Bhattacharya, G. A. Boorman, A. Boyles, B. U. Bradford, R. E. Bumgarner, P. R. Bushel, K. Chaturvedi, D. Choi D, et al., Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* **2**:351–356 (2005).

9. R. A. Irizarry, D. Warren, F. Spencer, I. F. Kim, S. Biswal, B. C. Frank, E. Gabrielson, J. G. Garcia, J. Geoghegan, G. Germino, et al., Multiple-laboratory comparison of microarray platforms. *Nat Methods* **2**:345–350 (2005).

10. J. E. Larkin, B. C. Frank, H. Gavras, R. Sultan, and J. Quackenbush, Independence and reproducibility across microarray platforms. *Nat Methods* **2**:337–344 (2005).

11. G. Sherlock, Of fish and chips. *Nat Methods* **2**:329–330 (2005).

12. S. E. Choe, M. Boutros, A. M. Michelson, G. M. Church, and M. S. Halfon, Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol* **6**:R16 (2005).

13. P. M. Holland, R. D. Abramson, R. Watson, and D. H. Gelfand, Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci USA* **88**:7276–7280 (1991).

14. K. J. Livak, S. J. Flood, J. Marmaro, W. Giusti, and K. Deetz, Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**:357–362 (1995).

15. C. A. Heid, J. Stevens, K. J. Livak, and P. M. Williams, Real time quantitative PCR. *Genome Res* **6**:986–994 (1996).

16. S. Tyagi and F. R. Kramer, Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14**:303–308 (1996).

17. D. Whitcombe, J. Theaker, S. P. Guy, T. Brown, and S. Little, Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* **17**:804–807 (1997).

18. V. V. Didenko, DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *Biotechniques* **31**:1106–1116, 1118, 1120–1121 (2001).

19. I. Nazarenko, B. Lowe, M. Darfler, P. Ikonomi, D. Schuster, and A. Rashtchian, Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. *Nucleic Acids Res* **30**:e37 (2002).

20. F. Pattyn, F. Speleman, A. De Paepe, and J. Vandesompele, RTPrimerDB: the real-time PCR primer and probe database. *Nucleic Acids Res* **31**:122–123 (2003).

21. X. Wang and B. Seed, A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res* **31**:e154 (2003).

22. S. A. Bustin, Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**(1):23–39 (2002).

23. J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**(7): RESEARCH0034 (2002).

24. S. J. Wall and D. R. Edwards, Quantitative reverse transcription-polymerase chain reaction (RT-PCR): a comparison of primer-dropping, competitive, and real-time RT-PCRs. *Anal Biochem* **300**(2):269–273 (2002).

25. M. W. Pfaffl, G. W. Horgan, and L. Dempfle, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* **30**(9): e36 (2002).

26. M. W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**(9): e45 (2001).

27. B. Weinstock-Guttman, D. Badgett, K. Patrick, L. Hartrich, R. Santos, D. Hall, M. Baier, J. Feichter, and M. Ramanathan, Genomic effects of IFN-beta in multiple sclerosis patients. *J Immunol* **171**(5):2694–2702 (2003).

28. I. Rioja, K. A. Bush, J. B. Buckton, M. C. Dickson, and P. F. Life, Joint cytokine quantification in two rodent arthritis models: kinetics of expression, correlation of mRNA and protein levels and response to prednisolone treatment. *Clin Exp Immunol* **137**(1):65–73 (2004).

29. H. Vermeer, B. I. Hendriks-Stegeman, B. van der Burg, S. C. van Buul-Offers, and M. Jansen, Glucocorticoid-induced increase in lymphocytic FKBP51 messenger ribonucleic acid expression: a potential marker for glucocorticoid sensitivity, potency, and bioavailability. *J Clin Endocrinol Metab* **88**(1):277–284 (2003).

30. H. J. Xie, U. Broberg, L. Griskevicius, S. Lundgren, S. Carlens, L. Meurling, C. Paul, A. Rane, and M. Hassan. Alteration of pharmacokinetics of cyclophosphamide and suppression of the cytochrome p450 genes by ciprofloxacin. *Bone Marrow Transplant* **31**(3):197–203 (2003).

31. J. Cummings, G. Boyd, J. S. Macpherson, H. Wolf, G. Smith, J. F. Smyth, and D. I. Jodrell, Factors influencing the cellular accumulation of SN-38 and camptothecin. *Cancer Chemother Pharmacol* **49**(3):194–200 (2002).

32. S. Cisternino, C. Mercier, F. Bourasset, F. Roux, and J. M. Scherrmann, Expression, up-regulation, and transport activity of the multidrug-resistance protein Abcg2 at the mouse blood-brain barrier. *Cancer Res* **64**(9):3296–3301 (2004).

33. G. U. Denk, C. J. Soroka, A. Mennone, H. Koepsell, U. Beuers, and J. L. Boyer, Down-regulation of the organic cation transporter 1 of rat liver in obstructive cholestasis. *Hepatology* **39**(5):1382–1389 (2004).

34. T. Schaarschmidt, J. Merkord, U. Adam, E. Schroeder, C. Kunert-Keil, B. Sperker, B. Drewelow, and R. Wacke, Expression of multidrug resistance proteins in rat and human chronic pancreatitis. *Pancreas* **28**(1):45–52 (2004).

35. G. Smith, R. S. Dawe, C. Clark, A. T. Evans, M. M. Comrie, C. R. Wolf, J. Ferguson, and S. H. Ibbotson, Quantitative real-time reverse transcription-polymerase chain reaction analysis of drug metabolizing and cytoprotective genes in psoriasis and regulation by ultraviolet radiation. *J Invest Dermatol* **121**(2):390–398 (2003).

36. S. Wilkening and A. Bader, Influence of culture time on the expression of drug-metabolizing enzymes in primary human hepatocytes and hepatoma cell line HepG2. *J Biochem Mol Toxicol* **17**(4):207–213 (2003).

37. T. C. DeLozier, C. C. Tsao, S. J. Coulter, J. Foley, J. A. Bradbury, D. C. Zeldin, and J. A. Goldstein, CYP2C44, a new murine CYP2C that metabolizes arachidonic acid to unique stereospecific products. *J Pharmacol Exp Ther* **310**(3):845–854 (2004).

38. T. Kogure, Y. Ueno, T. Iwasaki, and T. Shimosegawa, The efficacy of the combination therapy of 5-fluorouracil, cisplatin and leucovorin for hepatocellular carcinoma and its predictable factors. *Cancer Chemother Pharmacol* **53**(4):296–304 (2004).

39. S. Saito, A. Iida, A. Sekine, S. Kawauchi, S. Higuchi, C. Ogawa, and Y. Nakamura, Catalog of 680 variations among eight cytochrome p450 (CYP) genes, nine esterase genes, and two other genes in the Japanese population. *J Hum Genet* **48**(5):249–270 (2003).

40. M. Hiratsuka, K. Narahara, Y. Kishikawa, S. Ismail Hamdy, N. Endo, Y. Agatsuma, M. Matsuura, T. Inoue, Y. Tomioka, and M. Mizugaki, A simultaneous LightCycler detection assay for five genetic polymorphisms influencing drug sensitivity. *Clin Biochem* **35**(1):35–40 (2002).

41. A. Weise, S. Grundler, D. Zaumsegel, M. Klotzek, B. Grondahl, T. Forst, and A. Pfutzner, Development and evaluation of a rapid and reliable method for cytochrome P450 2C8 genotyping. *Clin Lab* **50**(3–4):141–148 (2004).

42. B. Eickhoff, B. Korn, M. Schick, A. Poustka, and J. van der Bosch, Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res* **27**(22): e33 (1999).

43. G. C. Tseng, M. K. Oh, L. Rohlin, J. C. Liao, and W. H. Wong, Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res* **29**(12):2549–2557 (2001).

44. B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**(2):185–193 (2003).

45. Z. Wu and R. A. Irizarry, Preprocessing of oligonucleotide array data. *Nat Biotechnol* **22**(6):656–658; author reply 658 (2004).

46. S. Draghici, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC Press, Boca Raton, FL., 2003.

47. Y. Zhao, M. C. Li, and R. Simon, An adaptive method for cDNA microarray normalization. *BMC Bioinformatics* **6**(1):28 (2005).

48. Y. Wang, J. Lu, R. Lee, Z. Gu, and R. Clarke, Iterative normalization of cDNA microarray data. *IEEE Trans Inf Technol Biomed* **6**(1):29–37 (2002).

49. X. Cui and G. A. Churchill, Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* **4**(4):210 (2003).

50. J. P. Brody, B. A. Williams, B. J. Wold, and S. R. Quake, Significance and statistical errors in the analysis of DNA microarray data. *Proc Natl Acad Sci USA* **99**(20):12975–12978 (2002).

51. P. Broberg, Statistical methods for ranking differentially expressed genes. *Genome Biol* **4**(6): R41 (2003).

52. X. L. Xu, J. M. Olson, and L. P. Zhao, A regression-based method to identify differentially expressed genes in microarray time course studies and its application in an inducible Huntington's disease transgenic model. *Hum Mol Genet* **11**(17):1977–1985 (2002).

53. Y. Zhao and W. Pan, Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics* **19**(9):1046–1054 (2003).

54. O. G. Troyanskaya, M. E. Garber, P. O. Brown, D. Botstein, and R. B. Altman, Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics* **18**(11):1454–1461 (2002).

55. M. K. Kerr and G. A. Churchill, Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc Natl Acad Sci USA* **98**(16):8961–8965 (2001).

56. M. K. Kerr and G. A. Churchill, Statistical design and the analysis of gene expression microarray data. *Genet Res* **77**(2):123–128 (2001).

57. M. K. Kerr, M. Martin, and G. A. Churchill, Analysis of variance for gene expression microarray data. *J Comput Biol* **7**(6):819–837 (2000).

58. J. Felsenstein, Confidence limits on phylogenies: an approach using bootstrap. *Evolution* **39**:783–791 (1985).

59. Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodological)* **57**:289–300 (1995).

60. Y. Benjamini and D. Yekutieli, The control of the false discovery rate in multiple tests under dependency. *Ann Stat* **29**:1165–1188 (2001).

61. J. D. Storey and R. Tibshirani, in *The Analysis of Gene Expression: Methods and Software*, G. Parmigiani, E. Garrett, R. A. Irizarry, and S. L. Zeger (Eds.). Springer, New York, 2003, pp. 272–290.

62. J. D. Storey and R. Tibshirani, Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**(16):9440–9445 (2003).

63. J. D. Storey, A direct approach to false discovery rates. *J Roy Stat Soc Ser B* **64**:479–498 (2002).

64. A. Reiner, D. Yekutieli, and Y. Benjamini, Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**(3):368–375 (2003).

65. P. H. Westfall and S. S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment,* Wiley, New York, 1993.

66. V. G. Tusher, R. Tibshirani, and G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**(9):5116–5121 (2001).

67. W. Pan, A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**(4):546–554 (2002).

68. G. Sherlock, Analysis of large-scale gene expression data. *Curr Opin Immunol* **12**(2):201–205 (2000).

69. A. Brazma and J. Vilo, Gene expression data analysis. *FEBS Lett* **480**(1):17–24 (2000).

70. A. Brazma and J. Vilo, Gene expression data analysis. *Microbes Infect* **3**(10):823–829 (2001).

71. S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, and I. Herskowitz, The transcriptional program of sporulation in budding yeast. *Science* **282**(5389):699–705 (1998).

72. P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**(12):3273–3297 (1998).

73. M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**(25):14863–14868 (1998).

74. B. Wu, T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, D. Ward, K. Williams, and H. Zhao, Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* **19**(13):1636–1643 (2003).

75. M. Q. Zhang, Discriminant analysis and its application in DNA sequence motif recognition. *Brief Bioinform* **1**(4):331–342 (2000).

76. T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**(10):906–914.

77. M. P. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, Jr., and D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* **97**(1):262–267 (2000).

78. C. H. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee, R. M. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, and T. Golub, Molecular classification of multiple tumor types. *Bioinformatics* **17**(Suppl 1):S316–S322 (2001).

79. M. Ringner, C. Peterson, and J. Khan, Analyzing array data using supervised methods. *Pharmacogenomics* **3**(3):403–415 (2002).

80. S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, Systematic determination of genetic network architecture. *Nat Genet* **22**(3):281–285 (1999).

81. P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, and T. R. Golub, Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* **96**(6):2907–2912 (1999).

82. A. Webb, *Statistical Pattern Recognition*, 2nd edition, Arnold Publishers/Oxford University Press, New York, 2002.

83. S. Raychaudhuri, J. M. Stuart, and R. B. Altman, Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*, pp. 455–466 (2000).

84. N. S. Holter, M. Mitra, A. Maritan, M. Cieplak, J. R. Banavar, and N. V. Fedoroff, Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci USA* **97**(15):8409–8414 (2000).

85. O. Alter, P. O. Brown, and D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* **97**(18):10101–10106 (2000).

86. K. Y. Yeung and W. L. Ruzzo, Principal component analysis for clustering gene expression data. *Bioinformatics* **17**(9):763–774 (2001).

87. P. D'Haeseleer, X. Wen, S. Fuhrman, and R. Somogyi, Linear modeling of mRNA expression levels during CNS development and injury. *Pac Symp Biocomput*, pp. 41–52 (1999).

88. B. Dysvik and I. Jonassen, J-Express: exploring gene expression data using Java. *Bioinformatics* **17**(4):369–370 (2001).

89. T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, 1995.

90. U. Fayyad, C. Reina, and P. S. Bradley, Initialization of iterative refinement clustering algorithms, in *Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, New York, 1998, pp. 194–198.

91. P. S. Bradley, U. Fayyad, and C. Reina, Scaling clustering algorithms to large databases, in *Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, New York, 1998, pp. 9–15.

92. L. J. Heyer, S. Kruglyak, and S. Yooseph, Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* **9**(11):1106–1115 (1999).

93. V. R. Iyer, M. B. Eisen, D. T. Ross, G. Schuler, T. Moore, J. C. Lee, J. M. Trent, L. M. Staudt, J. Hudson Jr., M. S. Boguski, D. Lashkari, D. Shalon, D. Botstein, and P. O. Brown, The transcriptional program in the response of human fibroblasts to serum. *Science* **283**(5398):83–87 (1999).

94. M. F. Ramoni, P. Sebastiani, and I. S. Kohane, Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA* **99**(14):9121–9126 (2002).

95. J. Vesanto and E. Alhoniemi, Clustering of the self-organizing map. *IEEE Transactions on Neural Networks* **11**:586–600 (2000).

96. D. A. Hosack, G. Dennis Jr., B. T. Sherman, H. C. Lane, and R. A. Lempicki, Identifying biological themes within lists of genes with EASE. *Genome Biol* **4**(10): R70 (2003).

97. G. Dennis Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**(5): P3 (2003).

98. P. Khatri, P. Bhavsar, G. Bawa, and S. Draghici, Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments, *Nucleic Acids Res* **32**(Web Server issue): W449–W456 (2004).

99. S. Draghici, P. Khatri, P. Bhavsar, A. Shah, S. A. Krawetz, and M. A. Tainsky, Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. *Nucleic Acids Res* **31**(13):3775–3781 (2003).

100. P. Khatri, S. Draghici, G. C. Ostermeier, and S. A. Krawetz, Profiling gene expression using onto-express. *Genomics* **79**(2):266–270 (2002).

101. M. Kanehisa and S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**(1):27–30 (2000).

102. M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**(Database issue): D277–D280 (2004).

103. H. Ogata, S. Goto, W. Fujibuchi, and M. Kanehisa, Computation with the KEGG pathway database. *Biosystems* **47**(1–2):119–128 (1998).

104. S. Goto, H. Bono, H. Ogata, W. Fujibuchi, T. Nishioka, K. Sato, and M. Kanehisa, Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput*, pp. 175–186 (1997).

105. S. W. Doniger, N. Salomonis, K. D. Dahlquist, K. Vranizan, S. C. Lawlor, and B. R. Conklin, MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol* **4**(1): R7 (2003).

106. K. D. Dahlquist, N. Salomonis, K. Vranizan, S. C. Lawlor, and B. R. Conklin, GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* **31**(1):19–20 (2002).

107. Anonymous. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. The IFNB Multiple Sclerosis Study Group. *Neurology* **43**(4):655–661 (1993).

108. Anonymous. PRISMS-4: Long-term efficacy of interferon-beta-1a in relapsing MS. *Neurology* **56**(12):1628–1636 (2001).

109. L. D. Jacobs, D. L. Cookfair, R. A. Rudick, R. M. Herndon, J. R. Richert, A. M. Salazar, J. S. Fischer, D. E. Goodkin, C. V. Granger, J. H. Simon, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Ann Neurol* **39**(3):285–294 (1996).

110. A. Gayo, L. Mozo, A. Suarez, A. Tunon, C. Lahoz, and C. Gutierrez, Interferon beta-1b treatment modulates TNFalpha and IFNgamma spontaneous gene expression in MS [see comments]. *Neurology* **52**(9):1764–1770 (1999).

111. R. A. Rudick, R. M. Ransohoff, J. C. Lee, R. Peppler, M. Yu, P. M. Mathisen, and V. K. Tuohy, In vivo effects of interferon beta-1a on immunosuppressive cytokines in multiple sclerosis. *Neurology* **50**(5):1294–1300 (1998). [Published erratum appears in *Neurology* **51**(1):332 (1998).]

112. R. M. Ransohoff, Cellular responses to interferons and other cytokines: the JAK-STAT paradigm. *N Engl J Med* **338**(9):616–618 (1998).

113. G. R. Stark, I. M. Kerr, B. R. Williams, R. H. Silverman, and R. D.Schreiber, How cells respond to interferons. *Annual Rev Biochem* **67**:227–264 (1998).

114. P. Chomczynski and N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**(1):156–159 (1987).

115. M. Reich, K. Ohm, M. Angelo, P. Tamayo, and J. P. Mesirov, GeneCluster 2.0: an advanced toolset for bioarray analysis. *Bioinformatics* **20**(11):1797–1798 (2004).

116. L. Hartrich, B. Weinstock-Guttman, D. Hall, D. Badgett, M. Baier, K. Patrick, J. Feichter, J. Hong, and M. Ramanathan, Dynamics of immune cell trafficking in interferon-beta treated multiple sclerosis patients. *J Neuroimmunol* **139**(1–2):84–92 (2003).

117. L. Zhang, A. Zhang, and M. Ramanathan, VizStruct: exploratory visualization for gene expression profiling. *Bioinformatics* **20**(1):85–92 (2004).

118. C. Tang, A. Zhang, and M. Ramanathan, ESPD: a pattern detection model underlying gene expression profiles. *Bioinformatics* **20**(6):829–838 (2004).

119. D. Jiang, J. Pei, M. Ramanathan, C. Tang, and A. Zhang, Mining coherent gene clusters from Gene-Sample-Time microarray data. *The 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (SIGKDD2004), Seattle, WA, 2004.

120. E. Q. Wang, W. I. Lee, D. Brazeau, and H. L. Fung, cDNA microarray analysis of vascular gene expression after nitric oxide donor infusions in rats: implications for nitrate tolerance mechanisms. *AAPS PharmSci* **4**(2):E10 (2002).

121. N. Friedman, Inferring cellular networks using probabilistic graphical models. *Science* **303**(5659):799–805 (2004).

122. T. Akutsu, S. Miyano, and S. Kuhara, Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J Comput Biol* **7**(3–4):331–343 (2000).

123. T. Akutsu, S. Miyano, and S. Kuhara, Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* **16**(8):727–734 (2000).

124. T. Akutsu, S. Miyano, and S. Kuhara, Algorithms for inferring qualitative models of biological networks. *Pac Symp Biocomput*, pp. 293–304 (2000).

125. T. Akutsu, S. Miyano, and S. Kuhara, Identification of genetic networks from a small number of gene expression patterns under the Boolean network mode. *Pac Symp Biocomput*, pp. 17–28 (1999).

126. T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano, A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. *Genome Inform Ser Workshop Genome Inform* **9**:151–160 (1998).

127. S. Liang, S. Fuhrman, and R. Somogyi, REVEAL, a general reverse engineering algorithm for inference of genetic network algorithms. *Pac Symp Biocomput*, pp. 18–29 (1998).

128. N. Friedman, M. Linial, I. Nachman, and D. Pe'er, Using Bayesian networks to analyze expression data. *J Comput Biol* **7**(3–4):601–620 (2000).

129. E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman, Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**(2):166–176 (2003).

130. T. Chen, H. L. He, and G. M. Church, Modeling gene expression with differential equations. *Pac Symposium Biocomputing*, pp. 29–40 (1999).

131. D. H. Irvine and M. A. Savageau, Efficient solution to nonlinear differential equations expressed in S-system canonical form. *SIAM J Numer Anal* **27**(3):704–735 (1990).

132. D. H. Irvine and M. A. Savageau, Network regulation of the immune response: modulation of suppressor lymphocytes by alternative signals including contrasuppression. *J Immunol* **134**(4):2117–2130 (1985).

133. D. H. Irvine and M. A. Savageau, Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J Immunol* **134**(4):2100–2116 (1985).

134. D. Tominaga and M. Okamoto, Design of a canonical model describing complex nonlinear dynamics. *Proc IFAC Int Conf CAB7 1998*, 1998, pp. 85–90.

135. W. J. Jusko and H. C. Ko, Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clin Pharmacol Ther* **56**(4):406–419 (1994).

136. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**(4):457–478 (1993).

137. R. Heinrich, B. G. Neel, and T. A. Rapoport, Mathematical models of protein kinase signal transduction. *Mol Cell* **9**(5):957–970 (2002).

138. Y. N. Sun and W. J. Jusko, Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* **87**(6):732–737 (1998).

139. M. Ramanathan, A dispersion model for cellular signal transduction cascades. *Pharm Res* **19**(10):1544–1548 (2002).

140. Z. X. Xu, Y. N. Sun, D. C. DuBois, R. R. Almon, and W. J. Jusko, Third-generation model for corticosteroid pharmacodynamics: roles of glucocorticoid receptor mRNA and tyrosine aminotransferase mRNA in rat liver. *J Pharmacokinet Biopharm* **23**(2):163–181 (1995).

141. S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold, Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**(3):1720–1730 (1999).

142. G. Chen, T. G. Gharib, C. C. Huang, J. M. Taylor, D. E. Misek, S. L. Kardia, T. J. Giordano, M. D. Iannettoni, M. B. Orringer, S. M. Hanash, and D. G. Beer, Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* **1**(4):304–313 (2002).

143. C. Ginestier, E. Charafe-Jauffret, F. Bertucci, F. Eisinger, J. Geneix, D. Bechlian, N. Conte, J. Adelaide, Y. Toiron, C. Nguyen, P. Viens, M-J Mozziconacci, R. Houlgatte, D. Birnbaum, and J. Jacquemier, Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol* **161**(4):1223–1233 (2002).

144. T. F. Orntoft, T. Thykjaer, F. M. Waldman, H. Wolf, and J. E. Celis, Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas. *Mol Cell Proteomics* **1**(1):37–45 (2002).

145. J. P. Dales, J. Plumas, F. Palmerin, E. Devilard, T. Defrance, A. Lajmanovich, V. Pradel, F. Birg, and L. Xerri, Correlation between apoptosis microarray gene expression profiling and histopathological lymph node lesions. *Mol Pathol* **54**(1):17–23 (2001).

146. J. L. Hargrove, M. G. Hulsey, F. H. Schmidt, and E. G. Beale, A computer program for modeling the kinetics of gene expression. *Biotechniques* **8**(6):654–659 (1990).

147. J. L. Hargrove and F. H. Schmidt, The role of mRNA and protein stability in gene expression. *Faseb J* **3**(12):2360–2370 (1989).

148. M. Ramanathan, R. D. MacGregor, and C. A. Hunt, Predictions of effect for intracellular antisense oligodeoxyribonucleotides from a kinetic model. *Antisense Res Dev* **3**(1):3–18 (1993).

149. D. Z. D'Argenio and A. Schlumitzky, *Users Guide to Release 4: Adapt II Pharmacokinetic/Pharmacodynamic Systems Analysis Software*, Biomedical Simulations Resource, University of Southern California, Los Angeles, CA, 1997.

150. T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**(5439):531–537 (1999).

151. A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D.Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**(6769):503–511 (2000).

152. I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O. P. Kallioniemi, A. Borg, and J. Trent, Gene-expression profiles in hereditary breast cancer. *N Engl J Med* **344**(8):539–548 (2001).

153. S. P. Bohen, O. G. Troyanskaya, O. Alter, R. Warnke, D. Botstein, P. O. Brown, and R. Levy, Variation in gene expression patterns in follicular lymphoma and the response to rituximab. *Proc Natl Acad Sci USA* **100**(4):1926–1930 (2003).

154. J. C. Chang, E. C. Wooten, A. Tsimelzon, S. G. Hilsenbeck, M. C. Gutierrez, R. Elledge, S. Mohsin, C. K. Osborne, G. C. Chamness, D. C. Allred, and P. O'Connell, Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* **362**(9381):362–369 (2003).

Pharmacogenomics and Pharmacokinetic/Pharmacodynamic Modeling

JIN Y. JIN and WILLIAM J. JUSKO

19.1 INTRODUCTION

Pharmacogenomics has focused on elucidating the genetic variation between individuals, its relationship with drug response and disease status, as well as how genes and proteins are regulated by exogenous compounds. The availability of sensitive measurement methods in pharmacogenomics provides an opportunity to examine key mechanisms affecting the diverse actions of drugs at the molecular level.

There are multiple potential sites in the pathway from DNA to protein that are available for regulation (Figure 19.1). All hormones have strong and diverse effects on gene expression and protein function. Steroid hormones (corticosteroids, estrogens, androgens) regulate transcription but not protein secretion. These are hydrophobic molecules that can easily diffuse through the cell membrane and bind with intracellular receptors that contain DNA-binding sites and can directly regulate transcription upon binding. On the other hand, peptide hormones (such as insulin and glucagon) regulate both transcription and protein activity. All peptide hormones along with catecholamines bind to receptors on the cell surface and initiate intracellular signal transduction events. Specific hormone-responsive DNA-binding proteins are produced to mediate transcriptional control. Second messengers such as cyclic AMP and calcium are generally involved in the transduction process. Furthermore, these hormones may also affect transcription, translation, and protein activity via phosphorylation. The endocrine system offers the additional complexity where one hormone may affect the secretion and function of other hormones.

Therefore, the pharmacogenomic (PG) study of hormonal drugs such as corticosteroids (CS) is of great importance to elucidate their mechanisms of action. Temporal patterns of dynamic changes in vivo will provide unique insights into the comprehensive regulation network.
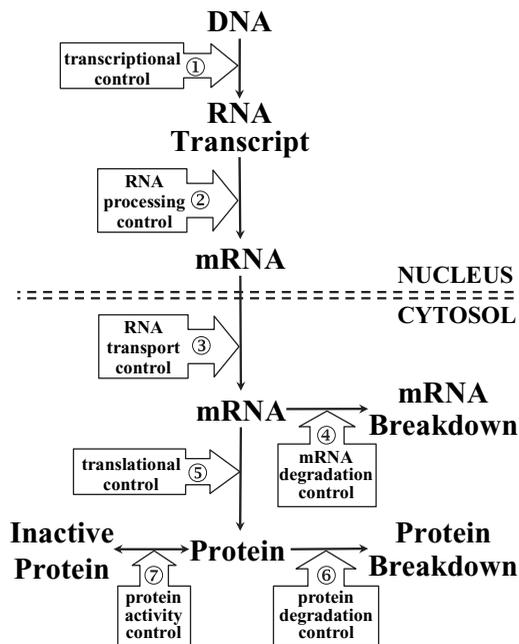
FIGURE 19.1   Steps in the pathway from DNA to protein that may be subject to regulation.

Advances in mechanism-based pharmacokinetic/pharmacodynamic (PK/PD) modeling have been steadily growing. There are a variety of mathematical models available for handling experimental data depending on the diverse mechanisms and rate-limiting processes that control drug effects (see review in Ref. 1). The multifaceted regulation of gene expression at multiple steps offers additional challenges in PK/PD modeling, which require integration of various "basic" dynamic models to describe the complex genomic system.

Understanding of the mechanism of drug action and measurements of the major contributing intermediate steps are essential for mechanistic modeling. Assay techniques to quantify message levels have evolved rapidly. Traditional methods such as Northern blot and RT-PCR only allow measurement of single genes. The development of microarrays enables simultaneous examination of thousands of genes and offers the opportunity to study the global picture of gene regulation. Due to the enormous amount of data obtained, there is a growing awareness of the need for development of mathematical models and other bioinformatic tools that will allow estimation of kinetic parameters that govern these biologic processes. PK/PD modeling could be used as a valuable tool to identify typical patterns and quantitatively describe the various mechanisms regulating genomic changes.

In this chapter, the complexity of PK/PD modeling in the field of pharmacogenomics is firstly demonstrated using selective genes. Results from gene arrays are then discussed to show the use of PK/PD modeling for studying thousands of genes at the same time. Corticosteroids were studied owing to their wide range of effects in pharmacogenomics and their various mechanisms of action in regulating gene

expression. The PD models provide quantitation of CS pharmacogenomics and exemplify the use of mathematical modeling to describe the molecular system.

19.2   MODELING OF SELECTIVE GENOMIC MARKERS

The cellular processes for CS pharmacogenomics are depicted in Figure 19.2. Unbound CSs in blood are moderately lipophilic and freely diffuse into the cytoplasm of liver cells. These steroids quickly bind to the cytosolic glucocorticoid receptor (GR) and cause activation of the receptor. This may lead to some rapid effects, such as cell trafficking, that seem not to depend on genomic mechanisms. The activated steroid–receptor complex may further translocate into the nucleus, where it can bind as a dimer to glucocorticoid responsive elements (GRE) in the target DNA and lead to the control of various genomic processes. The CS are known to cause homologous downregulation of their own receptors due to decreased transcription (2). After the transcriptional control of target genes, the steroid–receptor complexes in the nucleus may dissociate from the GRE, return to the cytosol, and are either degraded or recycled. In addition to the direct transcriptional regulation of target genes, CS may affect DNA transcription of functional biosignals/transcription factors (represented by BS) and cause changes in mRNA and protein concentrations of these regulators. Genes whose expression is controlled by these factors at transcriptional and/or post-transcriptional levels will thus be secondarily
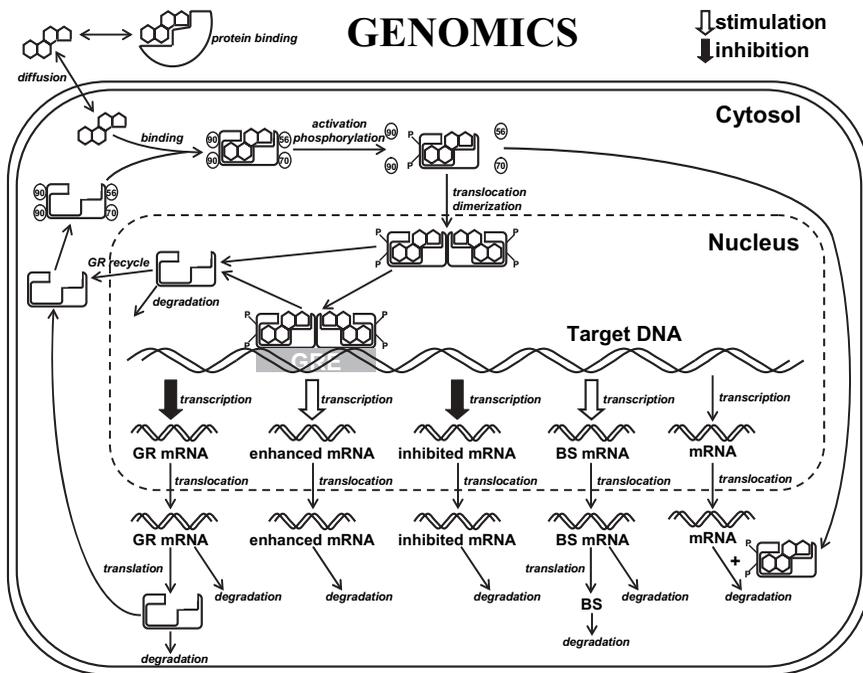
FIGURE 19.2   Schematic representation of diverse molecular mechanisms of CS action on gene expression in hepatocytes. GR, glucocorticoid receptor; GRE, glucocorticoid responsive element; BS, functional biosignal (such as transcription factors or hormones).

affected by CSs. The steroid–receptor complex may also directly or indirectly affect mRNA stability in cytosol.

19.2.1 Tyrosine Aminotransferase

Our laboratory has been involved extensively in modeling corticosteroid pharmacodynamics and related physiological systems. Modeling the genomic effects of CS has been an evolving process. Since 1986, a series of mechanism-based PK/PD models (Figure 19.3) have been proposed to describe the gene-mediated effects of steroids in terms of receptor downregulation and enzyme induction in rats (3–10). These models evolved over time with the addition of mRNA measurements and more extensive studies in terms of dosages used and duration of tissue sampling. Several basic PD mechanisms were integrated in these models including receptor binding, indirect response models for turnover steps, signal transduction, and tolerance development due to receptor downregulation. The tyrosine aminotransferase (TAT) enzyme in liver was selected as the PD marker because of the long history of use as an indicator of receptor/gene-mediated effects of CS in liver. The most recent "fifth-generation model" (Figure 19.3E) described the pharmacokinetics, GR mRNA repression, receptor dynamics, TAT mRNA induction, and its enzyme induction in liver upon acute dosing of methylprednisolone (MPL) in adrenalectomized (ADX) rats (7).

As shown in Figure 19.3E (inside the dotted square), the drug kinetics and receptor dynamics were modeled with the differential equations as follows:

$$C_{MPL} = C_1 \cdot e^{-\lambda_1 \cdot t} + C_2 \cdot e^{-\lambda_2 \cdot t} \tag{19.1}$$

$$\frac{dmRNA_R}{dt} = k_{s\_Rm} \cdot \left(1 - \frac{DR(N)}{IC_{50\_Rm} + DR(N)}\right) - k_{d\_Rm} \cdot mRNA_R \tag{19.2}$$

$$\frac{dR}{dt} = k_{s\_R} \cdot mRNA_R + R_f \cdot k_{re} \cdot DR(N) - k_{on} \cdot D \cdot R - k_{d\_R} \cdot R \tag{19.3}$$

$$\frac{dDR}{dt} = k_{on} \cdot D \cdot R - k_T \cdot DR \tag{19.4}$$

$$\frac{dDR(N)}{dt} = k_T \cdot DR - k_{re} \cdot DR(N) \tag{19.5}$$

where C_{MPL} represents the plasma concentration of MPL in ng/mL, D the plasma concentration of MPL in nmol/L, $mRNA_R$ the receptor mRNA, R the free cytosolic GR density, DR the cytosolic drug–receptor complex, and $DR(N)$ the drug–receptor complex in the nucleus. C_i and λ_i are coefficients for the intercepts and slopes of the PK profile. The rate constants in the equations include zero-order rate of GR mRNA synthesis (k_{s_Rm}); the first-order rates of GR mRNA degradation (k_{d_Rm}), receptor synthesis (k_{s_R}), and degradation (k_{d_R}), translocation of the drug–receptor complex into the nucleus (k_T), and the overall turnover of DR(N) to cytosol (k_{re}); as well as the second-order rate constant of drug–receptor association (k_{on}). In addition, IC_{50_Rm} is the concentration of DR(N) at which the synthesis rate of receptor mRNA drops to 50% of its baseline value, and R_f is the fraction of free receptor being recycled.
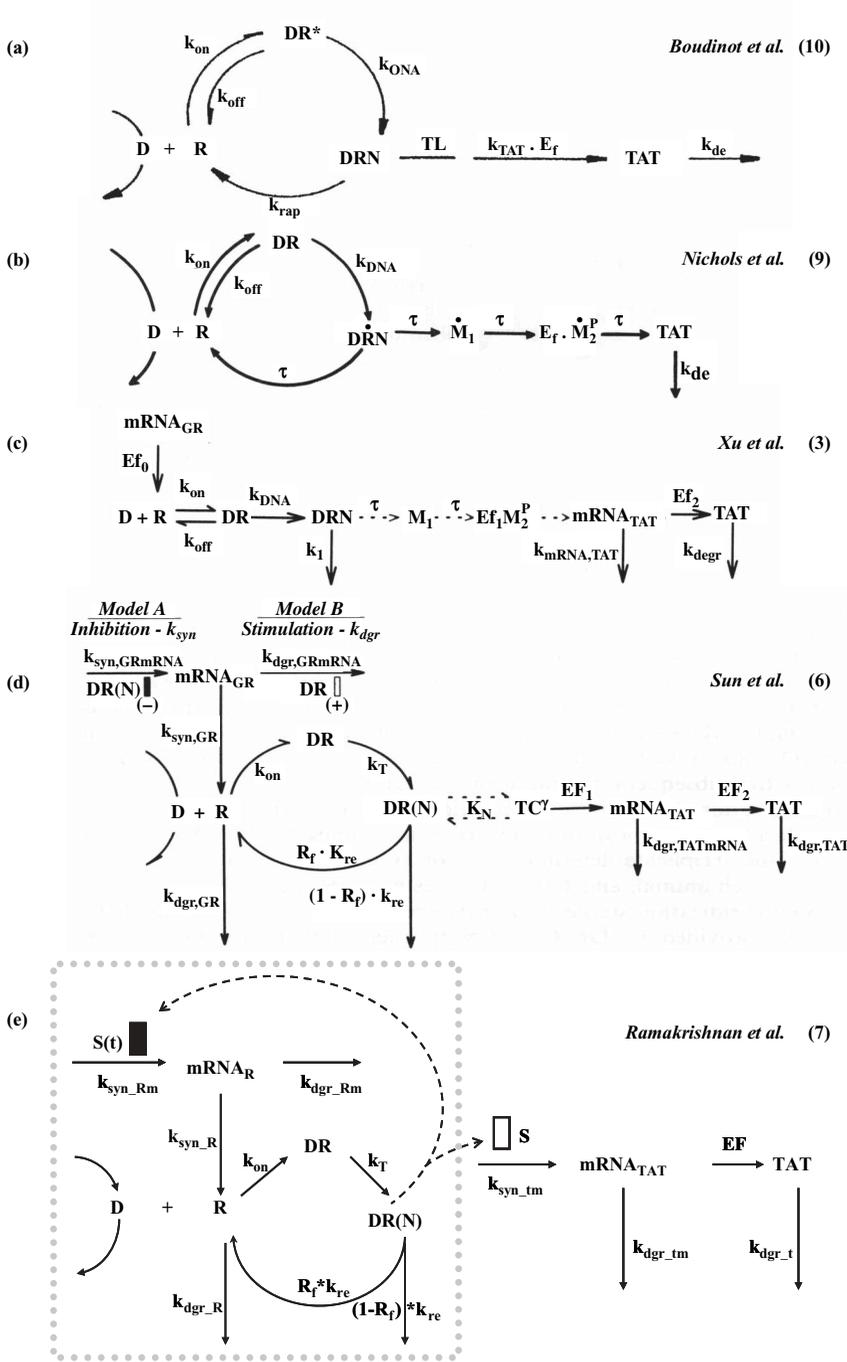
FIGURE 19.3 Proposed receptor/gene-mediated pharmacodynamic models for CS action: (A) first-generation model, (B) second-generation model, (C) third-generation model, (D) fourth-generation model, and (E) fifth-generation model.

The baselines were defined as

$$k_{d\_Rm} = \frac{k_{s\_Rm}}{mRNA_R^0} \tag{19.6}$$

$$k_{s\_R} = \left(\frac{R^0}{mRNA_R^0}\right) \cdot k_{d\_R} \tag{19.7}$$

where $mRNA_R^0$ and R^0 are the baseline values of receptor mRNA and free cytosolic GR density.

The TAT mRNA and activity were described by

$$\frac{dmRNA_{TAT}}{dt} = k_s^{TAm} \cdot \left(1 + S_M^{TAm} \cdot DR(N)\right) - k_d^{TAm} \cdot mRNA_{TAT} \tag{19.8}$$

$$\frac{dTAT}{dt} = k_s^{TA} \cdot (mRNA_{TAT})^{\gamma} - k_d^{TA} \cdot TAT \tag{19.9}$$

where $mRNA_{TAT}$ is the TAT message level in liver. The TAT mRNA is synthesized at a zero-order rate k_s^{TAm} and degraded at a first-order rate k_d^{TAm}. The drug–receptor binding complex in the nucleus, DR(N), stimulates TAT gene transcription with a linear efficiency factor S_M^{TAm}. The TAT is the hepatic TAT activity. The TAT enzyme was translated from its mRNA at the first-order rate k_s^{TA} with amplification factor γ and degraded at the first-order rate k_d^{TA}. The γ indicates that multiple copies of protein could be synthesized from a single mRNA transcript.

At time zero, the system baseline yields

$$k_s^{TAm} = k_d^{TAm} \cdot mRNA_{TAT}^0 \tag{19.10}$$

$$k_s^{TA} = \frac{k_d^{TA} \cdot TAT^0}{\left(mRNA_{TAT}^0\right)^{\gamma}} \tag{19.11}$$

where $mRNA^0_{TAT}$ and TAT^0 are the baseline values of TAT mRNA and activity.

Figure 19.4 shows the fitting results using the most recent "fifth-generation model," including the pharmacokinetics (A), GR mRNA repression (C), receptor dynamics (D), TAT mRNA induction (E), and its enzyme induction (F) in rat liver (6, 7). Also shown in Figure 19.4B is the simulated profile of drug–receptor complexes in the nucleus (DR(N)), which can act as the major driving force for genomic effects of CS. This model has been applied to simultaneously capture data profiles from other dose levels and infusions of methylprednisolone. Fitted model parameters have been published (7, 8).

19.2.2 Phosphoenolpyruvate Carboxykinase

Phosphoenolpyruvate carboxykinase (PEPCK) catalyzes the conversion of oxaloacetate to phosphoenolpyruvate, the rate-limiting step in hepatic and renal gluconeogenesis. Given its central role in gluconeogenesis, it is not surprising that PEPCK is tightly regulated by various hormones including CSs and glucagon (via cAMP).
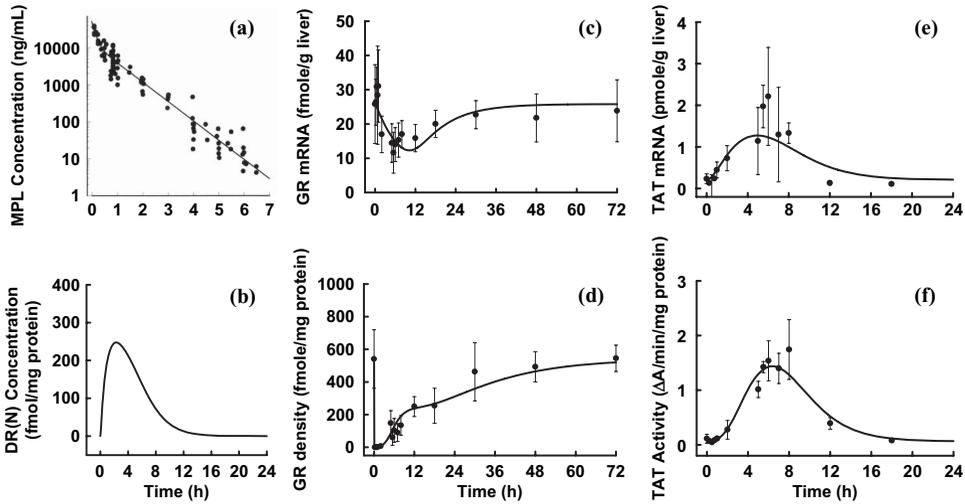
FIGURE 19.4 PK/PD modeling results using the fifth-generation model. Symbols and differential equations for the model are defined in Eqs. (19.1)–(19.11). Adrenalectomized rats received 50 mg/kg intravenous injection of MPL at 0 h. (A) Plasma MPL concentration; (B) simulated drug–receptor complex in hepatocyte nucleus (DR(N)); (C) GR mRNA level in liver; (D) free GR density in liver; (E) TAT mRNA level in liver; and (F) TAT activity in liver versus time. Solid circles are the observed data and bars are the standard deviations. Lines are model predictions. (Adapted from Refs. 6 and 7.)

The CS are known to enhance PEPCK gene transcription (11). The increased degradation rate of this message in liver after CS treatment has also been reported via an unknown mechanism (represented by hypothetical biosignal TC in Figure 19.5A) (12). The CS can also induce adenylyl cyclase and suppress phosphodiesterase based on similar transcriptional control (13). Adenylyl cyclase catalyzes the conversion of ATP to cAMP, which is degraded by phosphodiesterase (14). Alterations of these two enzymes by CS may result in increased cAMP levels in target cells. The translation of PEPCK mRNA to protein is known to be upregulated by cAMP (15). In addition, cAMP may reduce the stability of PEPCK enzyme (16). These actions result in the multifaceted regulation of PEPCK by multiple processes.

The drug kinetics and receptor dynamics were modeled as before (Section 19.2.1). As depicted in Figure 19.5A, a mechanistic model was proposed to describe the simultaneous actions of CS on PEPCK mRNA synthesis and degradation, induction of cAMP by CS, and translational stimulation by cAMP:

$$\frac{dTC}{dt} = k_1 \cdot (DR(N) - TC) \tag{19.12}$$

$$\frac{dmRNA_{PEPCK}}{dt} = k_s^{PM} \cdot \left(1 + S_s^{PM} \cdot DR(N)\right) - k_d^{Pm} \cdot \left(1 + S_d^{Pm} \cdot TC\right) \cdot mRNA_{PEPCK} \tag{19.13}$$

$$\frac{dcAMP}{dt} = k_s^C - k_d^C \cdot \left(1 - \frac{DR(N)}{IC_{50}^C + DR(N)}\right) \cdot cAMP \tag{19.14}$$
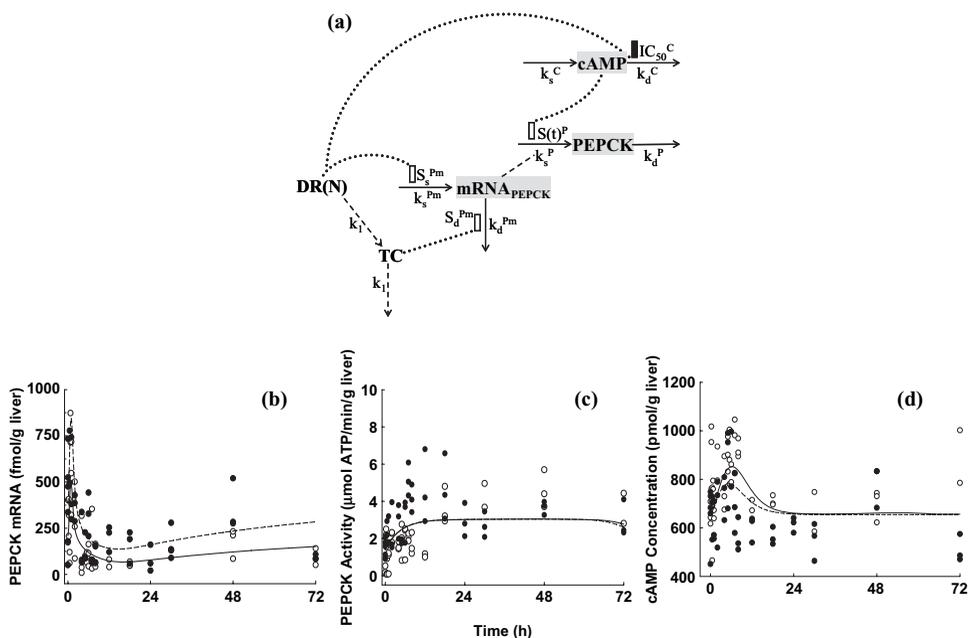
FIGURE 19.5 The proposed model (A) and fitting results for hepatic PEPCK mRNA (B), PEPCK activity (C), and cAMP levels (D) in rats receiving 10 (•) or 50 (○) mg/kg single MPL intravenous injection. Symbols and differential equations for the model are defined in Eqs. (19.12)–(19.18). The dotted lines and rectangles indicate stimulation (open bar) and inhibition (solid bar) of the various processes via indirect mechanisms. Lines in the graphs are results of the simultaneous fittings with Eqs. (19.12)–(19.15). Solid lines represent the high-dose group. Broken lines represent the low-dose group.

$$\frac{dPEPCK}{dt} = k_s^P \cdot S(t)^P \cdot (mRNA_{PEPCK})^\gamma - k_d^P \cdot PEPCK \qquad (19.15a)$$

$$S(t)^P = 1 + \frac{S_{max}^P \cdot (cAMP - cAMP^0)}{SC_{50}^P + (cAMP - cAMP^0)} \qquad (19.15b)$$

where *TC* is the concentration of the presumed biosignal responsible for the CS stimulation of PEPCK mRNA degradation. A linear transduction model (17) was used to describe this biosignal, which was generated from DR(N) at the first-order rate k_1. The $mRNA_{PEPCK}$ is the PEPCK message level in liver expressed as fmol/g liver. The stimulation of PEPCK mRNA synthesis rate k_s^{Pm} is dependent on DR(N) concentration with a linear efficiency factor S_s^{Pm}, and the stimulation of PEPCK mRNA degradation rate k_d^{Pm} is dependent on the transient TC with a linear efficiency factor S_d^{Pm}. The *cAMP* is the hepatic cAMP concentration in pmol/g liver. Endogenous cAMP is produced at a constant rate k_s^C. The IC_{50}^C represents the concentration of DR(N) producing 50% inhibition of cAMP degradation. The *PEPCK* enzyme was translated from its mRNA at the first-order rate k_s^P with amplification factor γ and degraded at the first-order rate k_d^P. The γ indicates that multiple copies of protein could be synthesized from a single mRNA transcript. S_{max}^P and SC_{50}^P

represent the maximum possible stimulation of k_s^P and the elevated cAMP required for half-maximal stimulation. The change of cAMP from its baseline value ($cAMP^0$) is used to drive this stimulation effect.

The baselines were defined as

$$k_s^{PM} = k_d^{Pm} \cdot mRNA_{PEPCK}^0 \tag{19.16}$$

$$k_s^C = k_d^C \cdot cAMP^0 \tag{19.17}$$

$$k_s^P = \frac{k_d^P \cdot PEPCK^0}{\left(mRNA_{PEPCK}^0\right)^\gamma} \tag{19.18}$$

where $cAMP^0, mRNA^0_{PEPCK}$, and $PEPCK^0$ are the baseline values of cAMP, PEPCK mRNA, and PEPCK activity.

Figure 19.5 shows the fitting results for PEPCK mRNA (B), PEPCK activity (C), and cAMP (D) in rat liver after a single injection of methylprednisolone. The acute tolerance/rebound phenomenon in PEPCK mRNA was nicely described by the dual action of CS on both gene transcription and degradation.

19.3 MODELING OF MICROARRAY PROFILES

Traditional assay techniques such as Northern blots and RT-PCR only allow measurement of a single or very small number of mRNA messages. In the past decade, the development of microarrays enables simultaneous examination of thousands of genes and offers the opportunity to study a more global picture of gene regulation. Due to the magnitude of data obtained by gene array studies, there is a growing awareness of the need for development of mathematical models and other bioinformatic tools that will allow estimation of kinetic parameters that govern the biologic processes.

Corticosteroid pharmacogenomics were recently studied using gene microarrays in rat liver (18, 19). In brief, methylprednisolone was administered intravenously at 50 mg/kg to adrenalectomized rats. Animals were sacrificed and livers excised at 17 time points over 72 hours. Four untreated rats were sacrificed at 0 h as controls. RNAs from individual livers were used to query Affymetrix GeneChips® (Affymetrix, Inc., Santa Clara, CA) which contain sequences for 8000 rat genes. Cluster analysis was performed using Affymetrix Microarray Suite 4.0® followed by GeneSpring 4.1® (Silicon Genetics, Redwood City, CA). Six temporal patterns consisting of 197 CS-responsive probes representing 143 genes were revealed from the cluster analysis. The whole data set is available online at `http://microarray.cnmcresearch.org/` (link *Programs* in *Genomic Applications*).

Based on our fifth-generation model of steroid pharmacokinetics/pharmacodynamics, mechanistic models were developed to describe the time pattern for each CS-responsive gene (18). The drug kinetics and receptor dynamics were modeled as before (Section 19.2.1). Based on the array of possible mechanisms (Figure 19.2), the following mathematical models were proposed to describe different gene expression patterns after MPL treatment in rat liver. In all of these models, target mRNA was assumed to be synthesized at zero-order rate k_{s_m} and degraded at first-order rate k_{d_m} without drug administration:

$$\frac{dmRNA}{dt} = k_{s\_M} - k_{d\_m} \cdot mRA \qquad (19.19)$$

The message level was assumed to be at steady-state at time zero (control animals), yielding the baseline equation

$$k_{s\_m} = k_{d\_m} \cdot mRNA^0 \qquad (19.20)$$

where $mRNA^0$ represents the baseline message level at time zero.

19.3.1 Simple-Regulated Genes

19.3.1.1 Induced Transcription

As depicted in Figure 19.6A, mRNA with induced production was described as follows:

$$\frac{dmRNA}{dt} = k_{s\_m} \cdot (1 + S \cdot DR(N)) - k_{d\_m} \cdot mRNA \qquad (19.21)$$

where the enhancement of transcription rate k_{s_m} is dependent on DR(N) concentration with a linear efficiency constant (S).

Figure 19.6B shows fitting of one selected gene using this model. These genes have patterns similar to tyrosine aminotransferase and could be well captured by our fifth-generation model. The mRNA degradation rate constant k_{d_m} represents the drug-independent property of the physiological system, and the linear stimulation factor S represents the drug-specific property of the message. Detailed results, descriptions, and discussion for this cluster and the other five clusters can be found in the original paper (18).
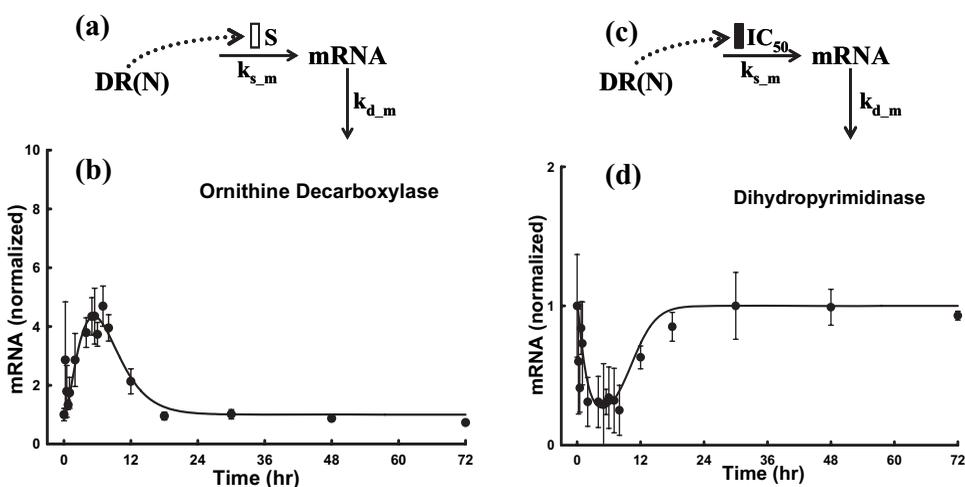
FIGURE 19.6 Proposed models and one representative fitting of induced and repressed genes. Symbols and differential equations for the models are defined in Eqs. (19.19)–(19.22). Solid circles are the mean gene array data and bars are the standard deviations. Solid lines are fittings with the proposed model for each individual gene.

19.3.1.2 Repressed Transcription

As depicted in Figure 19.6C, mRNA with repressed production was described as follows:

$$\frac{dmRNA}{dt} = k_{s\_m} \cdot \left(1 - \frac{DR(N)}{IC_{50} + DR(N)}\right) - k_{d\_m} \cdot mRNA \qquad (19.22)$$

where the inhibition of transcription rate k_{s_m} is dependent on DR(N) concentration, and IC_{50} represents the concentration of DR(N) at which the mRNA synthesis rate drops to 50% of its baseline value. Figure 19.6D shows fitting of one selected gene using this model.

19.3.2 Multifaceted-Regulated Genes

Some genes showed a more complex pattern with an initial decline followed by delayed increase, suggesting that two mechanisms might be involved. The primary and secondary drug effects in regulating the same biological system could be described by the generalized model (Figure 19.7). Drug (*Drug*) can produce its primary effect by altering the production (k_{in}) or disposition (k_{out}) of the biological marker (*Response*) via an indirect mechanism. Drug may also affect the level of an endogenous controlling factor (such as other hormones, cytokines, transcription factors), which is simply described by linear transduction. The rate constant (k) may reflect the major rate-limiting step producing this additional factor (*Biosignal*). This drug-altered biosignal regulates the same biological system via an indirect mechanism, causing the secondary effects in addition to the primary drug action. Drug and the biosignal may affect the system in the same fashion, which could represent multiple mechanisms of drug action in therapy. They may also produce counter-regulatory effects, which would be reflected as tolerance phenomena. Examples using specific combinations of drug and biosignal effects are shown in the following sections.

19.3.2.1 Repressed Transcription Plus Secondarily Induced Transcription by BS

As depicted in Figure 19.8A, mRNA with repressed then secondarily induced production was described as follows:
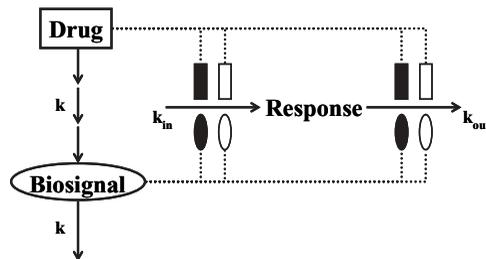
FIGURE 19.7   Model of primary and secondary drug effects. Symbols are defined in the text. The dotted lines and symbols indicate stimulation (open symbol) and inhibition (solid symbol) of the various processes via indirect mechanisms. The rectangles represent primary drug action. The ellipses represent secondary drug effects via biosignal action.
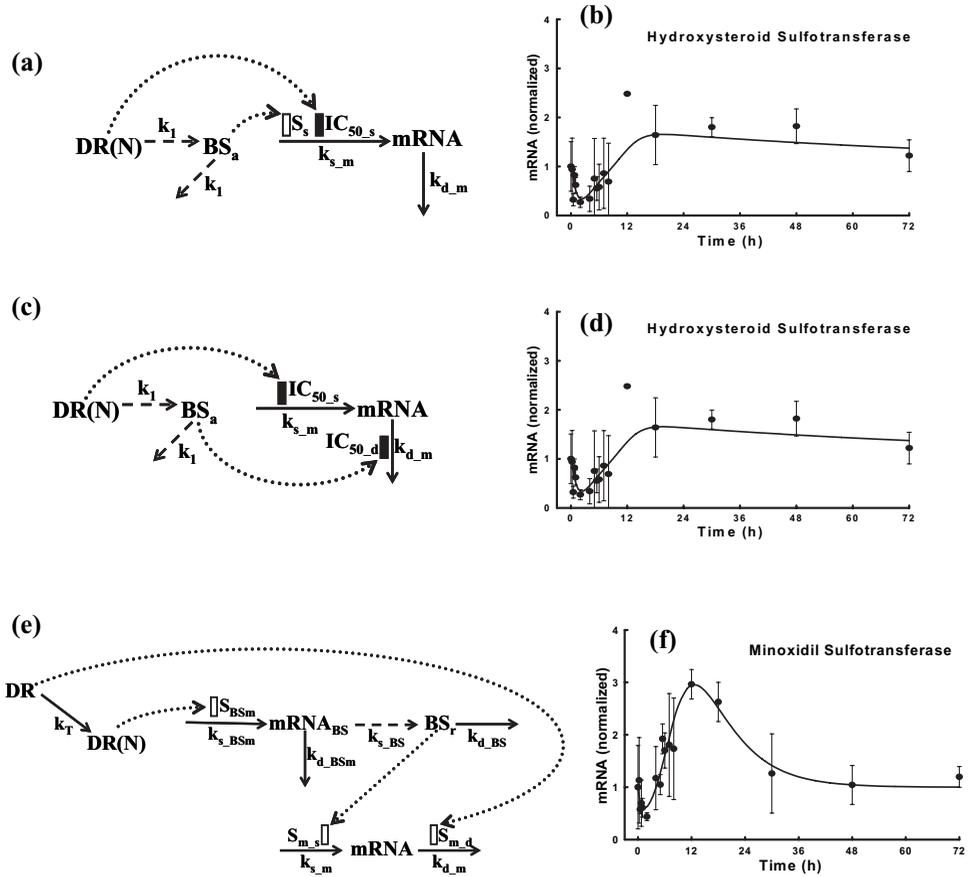
FIGURE 19.8 Proposed model and one representative fitting of multifaceted-regulated genes. Symbols and differential equations for the models are defined in Eqs. (19.23)–(19.32). Solid circles are the mean gene array data and bars are the standard deviations. Solid lines are fittings with the proposed model for each individual gene.

$$\frac{dBS_a}{dt} = k_1 \cdot (DR(N) - BS_a) \tag{19.23}$$

$$\frac{dmRNA}{dt} = k_{s\_m} \cdot \left(1 - \frac{DR(N)}{IC_{50\_s} + DR(N)} + S_s \cdot BS_a\right) - k_{d\_m} \cdot mRNA \tag{19.24}$$

where the intermediate regulator BS is described in a simplified fashion using a linear transduction model (17). In this model, BS_a represents the absolute change of regulator level from the control and this change is produced by $DR(N)$ via first-order rate (k_1). Drug ($DR(N)$) and the regulator (BS_a) both act on mRNA synthesis independently, characterized by a sigmoidal inhibition and a linear stimulation, respectively. The IC_{50_s} represents the concentration of DR(N) at which mRNA synthesis rate drops to 50% of its baseline value and S_s represents the efficiency of BS

stimulation on transcription. The hypothetical compartment (BS_a) is the simplified form of the primary-response product (BS_r) in the model in Section 19.3.2.3. This simplification allowed illustration of the underlying transduction process without assuming the behavior of the CS-enhanced biosignal. Figure 19.8B shows fitting of one selected gene using this model.

19.3.2.2 Repressed Transcription Plus Secondarily Repressed Degradation by BS

As depicted in Figure 19.8C, mRNA with repressed production and secondarily repressed degradation was described as follows:

$$\frac{dBS_a}{dt} = k_1 \cdot (DR(N) - BS_a) \tag{19.25}$$

$$\frac{dmRNA}{dt} = k_{s\_m} \cdot \left(1 - \frac{DR(N)}{IC_{50\_s} + DR(N)}\right) - k_{d\_m} \cdot \left(1 - \frac{BS_a}{IC_{50\_d} + BS_a}\right) \cdot mRNA \tag{19.26}$$

where BS_a represents the absolute change of the regulator level from the control and is characterized by a first-order rate constant (k_1). The inhibition of transcription rate k_{s_m} is dependent on DR(N) concentration, and IC_{50_s} represents the concentration of DR(N) at which mRNA synthesis rate drops to 50% of its baseline value. The inhibition of degradation rate k_{d_m} is dependent on absolute changes of BS, and IC_{50_d} represents the changes of BS_a at which mRNA degradation rate drops to 50% of its baseline value. The initial condition of Eq. (19.25) (BS_a^0) was fixed as 0. Figure 19.8D shows fitting of one selected gene using this model.

19.3.2.3 Induced mRNA Degradation in Cytosol Plus Secondarily Induced Transcription by BS

As depicted in Figure 19.8E, mRNA with DR-induced degradation and secondarily BS-induced production was described as follows:

$$\frac{dmRNA_{BS}}{dt} = k_{s\_BSm} \cdot (1 + S_{BSm} \cdot DR(N)) - k_{d\_BSm} \cdot mRNA_{BS} \tag{19.27}$$

$$\frac{dBS_r}{dt} = k_{s\_BS} \cdot mRNA_{BS} - k_{d\_BS} \cdot BS_r \tag{19.28}$$

$$\frac{dmRNA}{dt} = k_{s\_m} \cdot (1 + S_{m\_s} \cdot BS_r) - k_{d\_m} \cdot (1 + S_{m\_d} \cdot DR) \cdot mRNA \tag{19.29}$$

where symbols represent the message ($mRNA_{BS}$) and protein (BS_r) level of the intermediate regulator BS (both normalized as ratio to control). The extra rate constants in the equations include zero-order rate of BS mRNA synthesis (k_{s_BSm}); the first-order rates of BS mRNA degradation (k_{d_BSm}), translation to BS protein (k_{s_BS}), and protein degradation (k_{d_BS}). The stimulation of the BS transcription process k_{s_BSm} is dependent on DR(N) concentration with a linear efficiency constant (S_{BSm}). The stimulation of mRNA synthesis k_{s_m} is dependent on the relative changes of regulator BS with a linear efficiency constant (S_{m_s}). This stimulation is present even at baseline conditions. The mRNA degradation k_{d_m} in cytosol is regulated by DR concentration with a linear stimulation factor (S_{m_d}).

At time zero, the above equations yield the following baseline equations:

$$k_{s\_BSm} = k_{d\_BSm} \cdot mRNA_{BS}^0 \tag{19.30}$$

$$k_{s\_BS} = \left(\frac{BS_r^0}{mRNA_{BS}^0} \right) \cdot k_{d\_BS} \tag{19.31}$$

$$k_{s\_m} = \frac{k_{d\_m} \cdot mRNA^0}{1 + S_{m\_s} \cdot BS_r^0} \tag{19.32}$$

where $mRNA_{BS}^0$ and BS_r^0 are the baseline values of normalized BS mRNA and protein levels.

Figure 19.8F shows the fitting of one selected gene using this model. The initial decline was short-lived and was explained by a rapid stimulation of mRNA degradation in cytosol by steroid (DR). The predominant induction was assumed to be secondary to a CS-enhanced transcription factor (BS). The DR(N) enhanced the transcription of a BS gene in the nucleus, which translates to higher BS protein levels. The increase of BS leads to the delayed enhanced transcription of target genes. There is strong evidence in the literature that for at least two genes, arginase and carbamyl phosphate synthetase, the delayed induction was secondary to the primary CS-enhanced transcription factor C/EBP (20).

19.3.3   Other Issues

Our microarray data set contains substantial information about the mechanism and extent of CS effects on various genes. The small number of distinct temporal patterns indicates that a limited number of mechanisms may mediate CS pharmacogenomics. Of special note, several genes such as ornithine decarboxylase and hydroxysteroid sulfotransferase were represented by multiple different probes on the gene arrays. Reasonable concordance in profiles and dynamic parameters from multiple probes indicated a good degree of reproducibility in results.

The observed CS-responsive genes relate to a variety of biological processes. In this study, we analyzed genes individually. The present models serve to provide hypotheses on how mRNA expression is controlled by direct and secondary factors. These models sometimes confirm known mechanisms and sometimes are only possibilities that will need further exploration with specific studies. Text mining studies may identify biomedical literature findings that are relevant to the individual genes and, more importantly, the regulatory/functional relationship between the genes, the drug, and the tissues. Such efforts may lead to further integrated models incorporating multiple gene interregulations that will provide additional insights into signaling networks at molecular, cellular, and systemic levels.

Gene arrays are being increasingly used as probes for early steps in assessing primary and secondary effects of drugs in various tissues. Our studies show that such biomarkers obey biological rules of pharmacokinetics/pharmacodynamics and that a range of doses, sampling times, and study conditions will be needed to fully appreciate the relevance of altered gene expression profiles. Caution is needed in having too few time points as biphasic profiles occur frequently and certain time points will yield opposite conclusions.

19.4   MODELING METHODOLOGY

Data in these studies were generated from a so-called giant rat study in our laboratory. Animals were sacrificed to obtain serial blood and tissue samples. Each point represents the measurement from one individual rat and data from all these different rats were analyzed together to obtain a time profile as though it came from one "giant rat." A naive pooled data analysis approach was therefore employed for all model fittings using ADAPT II software (21). The maximum likelihood method was used with the variance model specified as $V(\sigma, \theta, t_i) = \sigma^2_1 Y(\theta, t_i)^{\sigma^2}$, where $V(\sigma, \theta, t_i)$ is the variance for the ith point, $Y(\theta, t_i)$ is the ith predicted value from the dynamic model, θ represents the estimated structural parameters, and σ_1 and σ_2 are the variance parameters that were estimated.

The comprehensive PD models were established by sequential fitting of dynamic markers in the biological cascade. Plasma drug concentrations over time were fitted then fixed to drive the dynamics in the following data analysis. Drug action was examined using different mathematical functions including linear and sigmoidal relationships with or without a Hill factor. When there were multiple mechanisms available, models were proposed based on each mechanism of action and fitted to the data. These models were compared based on visual inspection of curve fitting, estimator criterion value, sum of squared residuals, Akaike information criterion, Schwarz criterion, and confidence of parameter estimations. Once the optimal model was established and parameter estimates were obtained, they were fixed in the following data analysis. When there were multiple dosing regimens available, the model was fitted to data from all regimens simultaneously.

19.5   SUMMARY

There are multiple potential sites in the pathway from DNA to protein that are available for regulation. Drugs can regulate gene expression at various steps, alone or jointly with other transcription factors and/or hormones. The generalized mathematical models for receptor/gene/transduction dynamics facilitate the understanding of the global picture of drug actions and provide new insights for microarray data analysis. Pharmacogenomics provides an opportunity to examine multiple factors and mechanisms affecting the diverse molecular to whole-body actions of drugs. The PK/PD/PG models of the future will further integrate pharmacokinetics, molecular biology, and systems pharmacology.

ACKNOWLEDGMENTS

REFERENCES

1. D. E. Mager, E. Wyska, and W. J. Jusko, Diversity of mechanism-based pharmacodynamic models. *Drug Metab Dispos* **31**:510–518 (2003).

2. R. H. Oakley and J. A. Cidlowski, Homologous down regulation of the glucocorticoid receptor: the molecular machinery. *Crit Rev Eukaryot Gene Expr* **3**:63–88 (1993).

3. Z. X. Xu, Y. N. Sun, D. C. DuBois, R. R. Almon, and W. J. Jusko, Third-generation model for corticosteroid pharmacodynamics: roles of glucocorticoid receptor mRNA and tyrosine aminotransferase mRNA in rat liver. *J Pharmacokinet Biopharm* **23**:163–181 (1995).

4. Y. N. Sun, L. I. McKay, D. C. DuBois, W. J. Jusko, and R. R. Almon, Pharmacokinetic/pharmacodynamic models for corticosteroid receptor down-regulation and glutamine synthetase induction in rat skeletal muscle by a receptor/gene-mediated mechanism. *J Pharmacol Exp Ther* **288**:720–728 (1999).

5. Y. N. Sun, D. C. DuBois, R. R. Almon, N. A. Pyszczynski, and W. J. Jusko, Dose-dependence and repeated-dose studies for receptor/gene-mediated pharmacodynamics of methylprednisolone on glucocorticoid receptor down-regulation and tyrosine aminotransferase induction in rat liver. *J Pharmacokinet Biopharm* **26**:619–648 (1998).

6. Y. N. Sun, D. C. DuBois, R. R. Almon, and W. J. Jusko, Fourth-generation model for corticosteroid pharmacodynamics: a model for methylprednisolone effects on receptor/gene-mediated glucocorticoid receptor down-regulation and tyrosine aminotransferase induction in rat liver. *J Pharmacokinet Biopharm* **26**:289–317 (1998).

7. R. Ramakrishnan, D. C. DuBois, R. R. Almon, N. A. Pyszczynski, and W. J. Jusko, Fifth-generation model for corticosteroid pharmacodynamics: application to steady-state receptor down-regulation and enzyme induction patterns during seven-day continuous infusion of methylprednisolone in rats. *J Pharmacokinet Pharmacodyn* **29**:1–24 (2002).

8. R. Ramakrishnan, D. C. DuBois, R. R. Almon, N. A. Pyszczynski, and W. J. Jusko, Pharmacodynamics and pharmacogenomics of methylprednisolone during 7- day infusions in rats. *J Pharmacol Exp Ther* **300**:245–256 (2002).

9. A. I. Nichols, F. D. Boudinot, and W. J. Jusko, Second generation model for prednisolone pharmacodynamics in the rat. *J Pharmacokinet Biopharm* **17**:209–227 (1989).

10. F. D. Boudinot, R. D'Ambrosio, and W. J. Jusko, Receptor-mediated pharmacodynamics of prednisolone in the rat. *J Pharmacokinet Biopharm* **14**:469–493 (1986).

11. D. K. Granner, K. Sasaki, and D. Chu, Multihormonal regulation of phosphoenolpyruvate carboxykinase gene transcription. The dominant role of insulin. *Ann N Y Acad Sci* **478**:175–190 (1986).

12. W. Hoppner, W. Sussmuth, C. O'Brien, and H. J. Seitz, Cooperative effect of thyroid and glucocorticoid hormones on the induction of hepatic phosphoenolpyruvate carboxykinase in vivo and in cultured hepatocytes. *Eur J Biochem* **159**:399–405 (1986).

13. E. Baus, F. Van Laethem, F. Andris, S. Rolin, J. Urbain, and O. Leo, Dexamethasone increases intracellular cyclic AMP concentration in murine T lymphocyte cell lines. *Steroids* **66**:39–47 (2001).

14. R. W. Butcher, G. A. Robison, J. G. Hardman, and E. W. Sutherland, The role of cyclic AMP in hormone actions. *Adv Enzyme Regul* **6**:357–389 (1968).

15. J. M. Gunn, S. M. Tilghman, R. W. Hanson, L. Reshef, and F. J. Ballard, Effects of cyclic adenosine monophosphate, dexamethasone and insulin on phosphoenolpyruvate carboxykinase synthesis in Reuber H-35 hepatoma cells. *Biochemistry* **14**:2350–2357 (1975).

16. J. M. Gunn, F. J. Ballard, and R. W. Hanson, Influence of hormones and medium composition on the degradation of phosphoenolpyruvate carboxykinase (GTP) and total protein in Reuber H35 cells. *J Biol Chem* **251**:3586–3593 (1976).

17. Y. N. Sun and W. J. Jusko, Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* **87**:732–737 (1998).

18. J. Y. Jin, D. C. DuBois, R. R. Almon, N. A. Pyszczynski, and W. J. Jusko, Modeling of corticosteroid pharmacogenomics in rat liver using gene microarrays. *J Pharmacol Exp Ther* **307**:93–109 (2003).

19. R. R. Almon, D. C. DuBois, K. E. Pearson, D. A. Stephan, and W. J. Jusko, Gene arrays and temporal patterns of drug responses: corticosteroid effects in rat liver. *Functional Integrative Genomics* **3**:171–179 (2003).

20. T. Gotoh, S. Chowdhury, M. Takiguchi, and M. Mori, The glucocorticoid-responsive gene cascade. Activation of the rat arginase gene through induction of C/EBPbeta. *J Biol Chem* **272**:3694–3698 (1997).

21. D. Z. D'Argenio and A. Schumitzky, *ADAPT II User's Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software*. BioMedical Simulations Resource, Los Angeles, CA, 1997.

APPENDIX 19.1

The following example shows the modeling of a simply enhanced genomic marker after drug administration. The original study was performed by Ramakrishnan and colleagues (7). In this study, adrenalectomized rats received 50 mg/kg single IV injection of methylprednisolone. Rats were sacrificed by exsanguination under anesthesia at various time points after dosing. Four untreated rats were sacrificed at 0 h as controls. Blood and liver were collected at sacrifice and processed for the following assays. Plasma methylprednisolone concentrations, hepatic corticosteroid receptor mRNA, free receptor density, as well as tyrosine aminotransferase mRNA and activity were measured. The pharmacodynamic model was developed in the original article (7) and was described in detail in Section 19.2.1 (Eqs. (19.1)–(19.11)). The same model can also be used for microarray data in Section 19.3.1.1. As mentioned in Section 19.4, such comprehensive models were established by fitting a series of dynamic markers in the biological cascade sequentially. Due to space limitations, shown in this example is the last step of the model development, assuming drug kinetics and receptor dynamics have been established and fixed (Eqs. (19.1)–(19.7)). Tyrosine aminotransferase mRNA and activity are fitted as dynamic markers. This model was implemented in NONMEM Version 1.1 and the accompanying data were simulated using the previous parameter estimates (K_d^{TAm} = 0.38 h^{-1}, S_M^{TAm} = 0.29 L/nmol/mg protein, K_d^{TA} = 0.69 h^{-1}, γ = 1.8, $mRNA_{TAT}^0$ = 0.21 pmol/g, TAT^0 = 0.064 ΔA/mg protein) with ±10% random error.

Time (h)	TAT mRNA (pmol/g)	TAT Activity (ΔA/mg protein)
0	0.223	0.056
0.25	0.242	0.062
0.5	0.263	0.070
0.75	0.364	0.070
1	0.457	0.088
2	0.856	0.299
4	1.335	1.092
5	1.183	1.178
5.5	1.263	1.449
6	1.175	1.493
7	1.078	1.316
8	1.065	1.350
12	0.471	0.575
18	0.250	0.115
30	0.190	0.061
48	0.212	0.059
72	0.193	0.060

```
; Model for Simply Enhanced Genomic Marker
; TAT mRNA Induction by Corticosteroid Administration
; Developed by Ramakrishnan et al. JPP 29:1-24 (2002)
; Written by Jin and Jusko (2005)

$PROBLEM TAT mRNA induction

$INPUT ID, TIME, AMT, DV, EVID, MDV, CMT
$DATA eg.csv IGNORE=#

$SUBROUTINES ADVAN8 TOL=3

$MODEL
  COMP=(RECEPTOR)
  COMP=(DR)
  COMP=(DRN)
  COMP=(RMRNA)
  COMP=(TATMRNA)
  COMP=(TAT)

$PK
; PK MPL mol Concentration
  C1=39130
  C2=12670
  L1=7.54                                    ;lamda1
  L2=1.2                                     ;lamda2
  MPL=(C1*EXP(-L1*TIME)+C2*EXP(-L2*TIME))*1000/374.46
```

```
; Receptor Dynamics Parameters
  KSRM=2.9                                      ;Ksyn for RmRNA
  IC50=26.2                                     ;IC50 for RmRNA
  KON=0.00329
  KT=0.63
  KRE=0.57
  RF=0.49
  KDR=0.0572                                    ;Kdeg for R
  RM0=25.8                                      ;initial R mRNA
  REC0=540.7                                    ;initial R densigy
  KDRM=KSRM/RM0                                 ;Kdeg for RmRNA
  KSR=(REC0/RM0)*KDR                            ;Ksyn for R

; TATmRNA/TAT Dynamics Parameters
  KDTM=THETA(1)                                 ;Kdeg for TATmRNA
  STM=THETA(2)                                  ;S for TATmRNA
  KDT=THETA(3)                                  ;Kdeg for TAT
  GA=THETA(4)                                   ;gamma
  TAM0=THETA(5)                                 ;initial TAT mRNA
  TA0=THETA(6)                                  ;initial TAT
  KSTM=KDTM*TAM0                                ;Ksyn for TATmRNA
  KST=KDT*TA0/(TAM0**GA)                        ;Ksyn for TAT

  F1=REC0
  F2=0
  F3=0
  F4=RM0
  F5=TAM0
  F6=TA0

$DES
  DADT(1)= KSR*A(4)-KON*MPL*A(1)+KRE*RF*A(3)-KDR*A(1) ;R density
  DADT(2)= KON*MPL*A(1)-KT*A(2)                 ;DR
  DADT(3)= KT*A(2)-KRE*A(3)                     ;DR(N)
  DADT(4)= KSRM*(1-A(3)/(IC50+A(3)))-KDRM*A(4)  ;R mRNA
  DADT(5)= KSTM*(1+STM*A(3))-KDTM*A(5)          ;TAT mRNA
  DADT(6)= KST*(A(5)**GA)-KDT*A(6)              ;TAT

$ERROR
  IF (CMT.EQ.1) THEN
    Y=F+F*ERR(1)                               ;error for TAT mRNA
  ELSE
    Y=F+F*ERR(2)                               ;error for TAT
  ENDIF

$THETA (0, 0.38)                               ;KDTM
$THETA (0, 0.29)                               ;STM
$THETA (0, 0.69)                               ;KDT
```

```
$THETA (0, 1.8)                                        ;GA
$THETA (0, 0.223)                                      ;TAM0
$THETA (0, 0.056)                                      ;TA0

$OMEGA 0.04 0.04

$EST MAXEVAL=9999 PRINT=5 NOABORT
$COV

$TABLE ID AMT CMT TIME NOPRINT FILE=eg.tbl

$SCAT PRED VS DV UNIT
```

Empirical Pharmacokinetic/ Pharmacodynamic Models

JAMES A. UCHIZONO and JAMES R. LANE

20.1   INTRODUCTION

In 1937, Teorell's two articles (1, 2), "Kinetics of Distribution of Substances Administered to the Body," are generally credited as the origins of pharmacokinetics. Thus, his work launched an entire area of science that ultimately deals with the quantitative aspects of designing rational dosing regimens that have the highest probability of achieving the targeted position on the response surface (3, 4). Frequently, it is said that "pharmacokinetics is what the body does to the drug, and pharmacodynamics is what the drug does to the body." The inextricable link between pharmacokinetics and pharmacodynamics has led to the frequently used pharmacokinetic/pharmacodynamic (PK/PD) descriptor in the literature. Connecting pharmacokinetics to pharmacodynamics is often accomplished through a hypothetical effect (or biophase) compartment described by "link" models (5–7). The combination of a pharmacokinetic, a link, and a PD model will mathematically describe the relationship between dose (input) and effect or response (output) (see Figure 20.1).

Figure 20.1 shows the PK model translating dose into a plasma concentration (C_p); the link model mapping C_p into the drug concentration at the effect site (C_e); and finally, the PD model translates C_e into effect. For most drugs, C_p at steady-state is in one-to-one correspondence with effect, which leads to fairly simple mathematical models that can relate dose to effect. In cases where this one-to-one correspondence does not exist or in cases where the link and/or PD kinetics are slow relative to changes in dose, the time-varying nature of the system is usually modeled as a biophase or a biosensing or a transduction kinetics problem. This chapter covers empirical, direct PK/PD models and two models with time- or state-varying parameters, one from pharmacokinetics (time-varying, clearance) and one from pharmacodynamics (state-varying, drug tolerance).
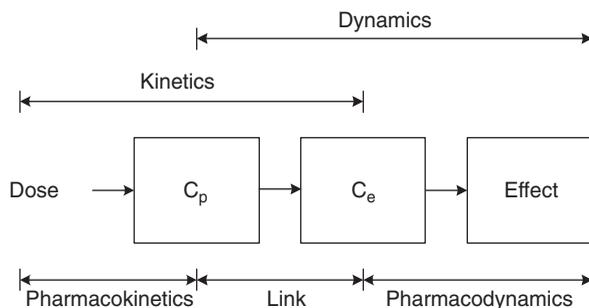
FIGURE 20.1 The relationship between pharmacokinetics, link model, and pharmacodynamics.
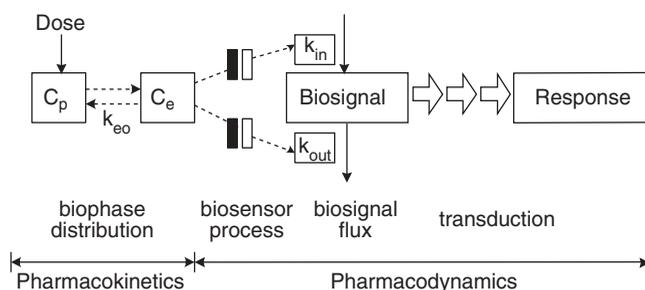
FIGURE 20.2 The different processes within pharmacodynamics. The biophase kinetics account for distributional kinetics between the plasma and effect site. The biosensor kinetics describe the dynamic behavior between drug and biosensing machinery. The biosignal is the first messenger in the transduction process that converts the signal to secondary messengers or the observed response. (Adapted from Ref. 19.)

For at least the last 30 years several types of PD models (6, 8–17) have been coupled to PK models to give a full description from input (i.e., dose) to effect (i.e., observed pharmacological response). Several classifications have arisen to describe the PD aspect: *empirical*, *direct*, and *indirect*. This chapter focuses on empirical and direct models, which are sometimes used synonymously in pharmacokinetics/pharmacodynamics. Although "empirical" models have parameters that tend not to be based on mechanistic underpinnings, direct models have some mechanistic reality, as well as empirical attributes (18).

In Figure 20.1 the PD model can be broken down further (see Figure 20.2) to describe a more mechanistic view. In Figure 20.2 the kinetics of drug action can be characterized by a combination of four kinetic processes: input and disposition kinetics (20–22), biophase or effect site equilibration kinetics (usually does not impact the PK disposition kinetics) (6, 23–26), biosensor/biosignal kinetics (e.g., drug–receptor binding kinetics) (27–32), and transduction kinetics (e.g., second messenger cascades) (19, 33–35). The first two are described in the PK model; the last two are described in the PD model; the biophase or effect site is part of both and connects the pharmacokinetics to the pharmacodynamics. Most direct models assume the latter two are instantaneous or rapid compared to all other kinetic drug action processes; this assumption is a distinguishing characteristic between direct

and indirect models. While the distinction between the two is somewhat arbitrary, in that nearly all mechanisms of drug action can be modeled as *indirect*, the *direct* or *empirical* models continue to have practical utility. In the next section, direct models ranging from semimechanistic (E_{max} model) to mostly empirical (cubic splines) PK/PD models are covered.

20.2 DIRECT EMPIRICAL MODELS

Direct models are the most prevalent PD models in the literature. In their most basic form, direct models associate drug concentration or "intensity" directly to the measured effect using stationary (time-invariant) parameters, meaning changes in blood/plasma drug concentrations are instantaneously realized in the observed effect. While the mathematical formulation of direct models can easily be distinguished from indirect models, direct models can also be viewed as a special subset of indirect models. Direct PD models have their roots in the work of Levy (36, 37). Levy proposed the following two models (Eqs. (20.1) and (20.2)) to describe a direct relationship between drug plasma concentration (C_p) and effect (E)

$$E = E_0 \pm S \times C_p \tag{20.1}$$

$$E = E_0 \pm m \times \log(C_p) \tag{20.2}$$

where E is effect, C_p is drug plasma concentration, E_0 is the baseline effect, S is the slope of the linear relationship, and m is the slope of the log-linear relationship. Since most drugs do have a specific concentration range in which effect or response is directly proportional to drug concentration, these models do capture some observed concentration–effect data. When drug concentrations are relatively low, Eq. (20.1) works well. As drug concentration is increased, the concentration–effect relationship is better described by the log-linear Eq. (20.2), which occurs at concentrations that produce effects in the 20–80% range of maximal effect. The main weakness of these models is their unrealistic extrapolation maximum—infinity; the maximum effect is not limited and will continue to increase with increasing concentration. This deficiency is removed in the E_{max} model.

The well known E_{max} model, Eq. (20.3), has its mechanistic basis in the *law of mass action* (16, 29–31, 38),

$$E = \frac{E_{max} \times C_p}{EC_{50} + C_p} \tag{20.3}$$

$$E = E_0 \pm \frac{E_{max} \times C_p}{EC_{50} + C_p} \tag{20.4}$$

where E is the observed effect, E_{max} is the maximum effect, EC_{50} is the drug concentration at which 50% E_{max} effect is observed, and C_p is the drug concentration. E_0, the baseline effect, is added in Eq. (20.4) to further capture biological reality. Just as in Eqs. (20.1) and (20.2), the ± symbol is used to indicate that the drug can increase or decrease the observed effect, depending on the effect system being

studied. Figure 20.3 shows a plot of effect versus concentration for this model. Since biological systems have limited resources (e.g., receptor protein), the expectation that the effect should plateau is met with the E_{max} model. Additionally, this model, given by Eq. (20.4), encapsulates the behavior modeled in Eqs. (20.1) and (20.2). When $C_p \ll EC_{50}$, the denominator essentially becomes EC_{50}, shown in Eq. (20.5):

$$E \approx E_0 \pm \frac{E_{max} \times C_p}{EC_{50}} = E_0 \pm S \times C_p \tag{20.5}$$

where $S = E_{max}/EC_{50}$ and the effect is linear with concentration similar to Eq. (20.1). When $C_p \gg EC_{50}$, the denominator becomes C_p, shown in Eq. (20.6):

$$E \approx E_0 \pm \frac{E_{max} \times C_p}{C_p} = E_0 \pm E_{max} \tag{20.6}$$

where the response is independent of C_p. When C_p is between these two limiting cases (effect is between 20% and 80% of E_{max}), the relationship between C_p and effect is log-linear, similar to Eq. (20.2). With the addition of an exponent, γ (analogous to m in Eq. (20.2)), to the E_{max} model, the slope of the log-linear region can be controlled; this model is known as the *sigmoidal* E_{max} model (Eq. (20.7)) shown in Figure 20.4.

$$E = E_0 \pm \frac{E_{max} \times C_p^{\gamma}}{EC_{50}^{\gamma} + C_p^{\gamma}} \tag{20.7}$$

The γ parameter does not need to be placed on the EC_{50}, but doing so simplifies maintaining integrity between the units of EC_{50} and C_p. Since a majority of drugs
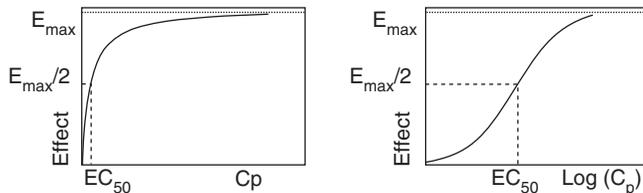
FIGURE 20.3   The left panel shows a Cartesian plot for an E_{max} model. EC_{50} is the C_p concentration that produces one-half the maximal effect (E_{max}). The right panel is the same data plotted with a logarithmic abscissa.
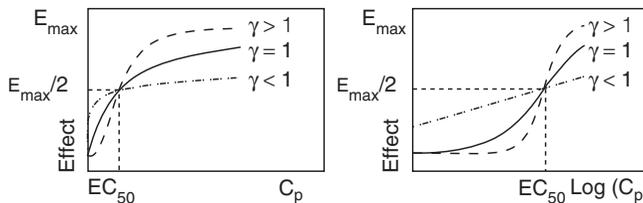
FIGURE 20.4   The left panel shows a Cartesian plot for a sigmoidal E_{max} model with three different values of γ. The right panel is the same data plotted with a logarithmic abscissa.

work through receptor or receptor-like mediated processes and this direct model is based on the law of mass action, the sigmoidal E_{max} has tremendous utility. At low concentrations it predicts a linear concentration–effect relationship, at higher concentrations it predicts a log-linear relationship, and at very high concentrations it predicts a constant effect independent of concentration. From a fitting perspective, unless E_{max} is known or operationally achieved experimentally, Eqs. (20.4) and (20.7) should be fit carefully to avoid identifiable issues (39). Several researchers have developed strategies that increase the reliability of these parameter estimates (40–42). The utility and intuitive nature of this model have made it the workhorse of pharmacodynamics.

Another class of empirical models, which are nearly void of any mechanistic basis, used in complex dose–concentration and concentration–effect PK/PD models are splines (43–46). Depending on how splines are used (explicitly mapping C_p into effect—i.e., effect = $f(C_p)$, where $f(C_p)$ is the spline function), they can also be considered direct PD models. *Empirical* models, as the name implies, emphasize phenomenological relationships. Toward one extreme these models fit data without any regard to underlying mechanistic underpinnings. Polynomial fitting and spline fitting fall into this category (47–49). For example, for n sets of data points (x, y), a polynomial, $P(x)$, of order $n - 1$ (Eq. (20.8)) is guaranteed to exist, given no parameter constraints, that will pass through all (x, y) data.

$$\hat{y} = P(x) = \beta_0 + \beta_1 x + \cdots + \beta_{n-1} x^{n-1} \tag{20.8}$$

In Eq. (20.8) β's are parameter constants and x is the independent or predictor variable. The advantages of this model are (a) it fits all data points (x, y), (b) $P(x)$ is easy to differentiate and integrate, and (c) interpolation of y along the interval between the smallest value and the largest values of x (i.e., $Min(x) \leq x \leq Max(x)$) tends to be quite good. The disadvantages to this model are (a) it assumes no error in data; (b) as n gets large, typical objective functions are overly sensitive to contributions by high-order terms—leading to fitting difficulties; (c) there is generally poor extrapolation of y along the intervals outside $Min(x) \leq x \leq Max(x)$; and (d) parameters have little to no physical meaning or relationship to the system being studied.

Another polynomial related approach utilizes cubic splines. While these splines are not plagued by high-order term sensitivity, they too do not provide parameters with physical meaning. There are numerous classes and subclasses of splines, each with unique properties (47–49); one commonly used spline is the cubic spline.

Simple cubic splines, $S_1(x)$ and $S_2(x)$, Eqs. (20.9), are piecewise continuous third-order polynomials governed by various constraints (depending on the type of cubic spline—e.g., *free* or *clamped*) that link three or more contiguous (x, y) points together. Spline segments are sequentially added until all n (x, y) data points along the interval, $Min(x) \leq x \leq Max(x)$, have been included to form $S(x)$.

$$\begin{aligned} S_1(x) &= \beta_{11} + \beta_{12} x + \beta_{13} x^2 + \beta_{14} x^3 \\ S_2(x) &= \beta_{21} + \beta_{22} x + \beta_{23} x^2 + \beta_{24} x^3 \\ &\vdots \qquad\qquad \vdots \\ S_k(x) &= \beta_{k1} + \beta_{k2} x + \beta_{k3} x^2 + \beta_{k4} x^3 \end{aligned} \tag{20.9}$$

where $k = n - 1$, β's are parameters, and $S(x)$ is given by Eq. (20.10):

$$S(x) = S_1(x)\big|_{x_1}^{x_2}; S_2(x)\big|_{x_2}^{x_3}; \cdots; S_k(x)\big|_{x_k}^{x_n} \tag{20.10}$$

Although a simpler second-order or quadratic spline of the form $S(x) = \beta_1 + \beta_2 x + \beta_3 x^2$ does guarantee a fit along the $\mathrm{Min}(x) \le x \le \mathrm{Max}(x)$ interval, it does not guarantee differentiability at the splines' exterior points; thus, a smooth, continuous fit cannot be guaranteed. The additional degree of freedom utilized in cubic splines allows users to define first and second derivative conditions at interior and exterior points (Figure 20.5). For example, for any three sequential $((x_1, y_1), (x_2, y_2), (x_3, y_3))$ pairs, eight conditions must be defined for the two connecting spline segments (S_1 and S_2) to ensure that all eight parameters (β's) can be determined. Conditions 1 through 4 ensure that each spline passes through its two endpoints (Eq. (20.11)):

$$
\begin{aligned}
\text{Condition 1} \quad & S_1(x_1) = \beta_{11} + \beta_{12}x_1 + \beta_{13}x_1^2 + \beta_{14}x_1^3 \\
\text{Condition 2} \quad & S_1(x_2) = \beta_{11} + \beta_{12}x_2 + \beta_{13}x_2^2 + \beta_{14}x_2^3 \\
\text{Condition 3} \quad & S_2(x_2) = \beta_{21} + \beta_{22}x_2 + \beta_{23}x_2^2 + \beta_{24}x_2^3 \\
\text{Condition 4} \quad & S_2(x_3) = \beta_{21} + \beta_{22}x_2 + \beta_{23}x_2^2 + \beta_{24}x_2^3
\end{aligned} \tag{20.11}
$$

Since conditions 1 to 4 only provide four equations to solve for eight unknowns, derivative and boundary conditions are utilized to provide the remaining four necessary equations. Conditions 5 and 6 match first (Eq. (20.12)) and second derivatives (Eq. (20.13)) at the nodes where two segments meet to ensure a smooth transition from S_i to S_{i+1}.

$$S_1'(x_2) = S_2'(x_2)$$

$$\text{Condition 5} \quad \beta_{12} + 2\beta_{13}x_2 + 3\beta_{14}x_2^2 = \beta_{22} + 2\beta_{23}x_2 + 3\beta_{24}x_2^2 \tag{20.12}$$

$$S_1''(x_2) = S_2''(x_2)$$

$$\text{Condition 6} \quad 2\beta_{13} + 6\beta_{14}x_2 = 2\beta_{23} + 6\beta_{24}x_2 \tag{20.13}$$
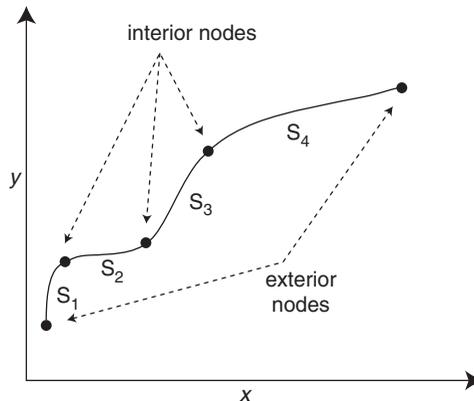
FIGURE 20.5   An example plot showing four spline segments connecting three interior nodes and two exterior nodes.

The last two conditions are for the exterior nodes and determine whether the cubic spline is natural/free or clamped. For natural splines conditions 7 and 8 are given by Eq. (20.14):

$$\text{Condition 7a} \quad S_1''(x_1) = 2\beta_{13} + 6\beta_{14}x_1 = 0$$
$$\text{Condition 8a} \quad S_2''(x_3) = 2\beta_{23} + 6\beta_{24}x_3 = 0 \tag{20.14}$$

In clamped cubic splines, conditions 7 and 8 define the first derivatives at the exterior nodes, which are constrained to fixed values, $f'(x_1)$ and $f'(x_3)$ (Eq. (20.15)), that are known beforehand. In the absence of precise first derivative information at the exterior nodes, one generally chooses Eq. (20.15) or the natural spline conditions:

$$\text{Condition 7b} \quad S_1'(x_1) = f'(x_1)$$
$$\text{Condition 8b} \quad S_2'(x_3) = f'(x_3) \tag{20.15}$$

Given the eight conditions, all eight β's can be determined for any two connecting segments; this process is continued until S_k has been determined. It must be noted that spline parameters do not have physical meaning and should *not* be used for extrapolation; they provide excellent interpolation. Polynomial fitting and spline fitting are classic examples of nearly pure empirical models.

20.3 PHARMACOKINETIC/PHARMACODYNAMIC LINK MODELS

In the previously mentioned direct and empirical models, it was assumed that the biophase kinetics, biosensing/biosignal kinetics, and transduction kinetics were all rapid compared to changes in C_p was made. However, there usually exists a time delay between giving the dose and the onset of effect (see Figure 20.6). Some drug
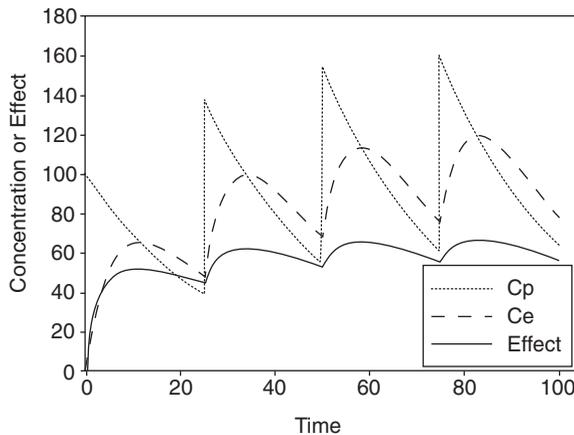
FIGURE 20.6 This plot shows three time profiles (C_p, C_e, *Effect*) produced in response to repeated IV bolus dosing every 25 time units. The time lag between $C_{p,\max}$ and $C_{e,\max}$ in this model is solely governed by the rate constant k_{e0}. Since C_e drives the effect, *Effect* also lags behind C_p.

classes (e.g., antidepressants) experience considerable delays before reaching full therapeutic effect. This section focuses on one strategy, biophase or effect compartment modeling, to account for the delay between dosing changes and their subsequent effect changes.

Although in 1955 Furchgott (5) did introduce the term biophase and biophase kinetics as a means to differentiate the actions of various agonists on vascular smooth muscle, in 1968 Segre (7) created a biophase compartmental model with first-order kinetics and the idea of a hypothetical effect compartment was born. In 1978 Dalhstrom et al. (26) modeled the relationship between morphine and analgesia using the sum of two effect compartments in a linear combination to produce the overall effect. In 1979 Sheiner et al. (6) linked d-tubocurarine pharmacokinetics (multicompartment) to its pharmacodynamics using a single hypothetical effect compartment (A_e or amount of drug in the effect compartment), which utilized a fractional, direct sigmoidal E_{max} model between A_e and effect shown in Eq. (20.16) (see Figure 20.7):

$$E = \frac{A_e^\gamma}{A_e^\gamma + A_e(50)^\gamma} \tag{20.16}$$
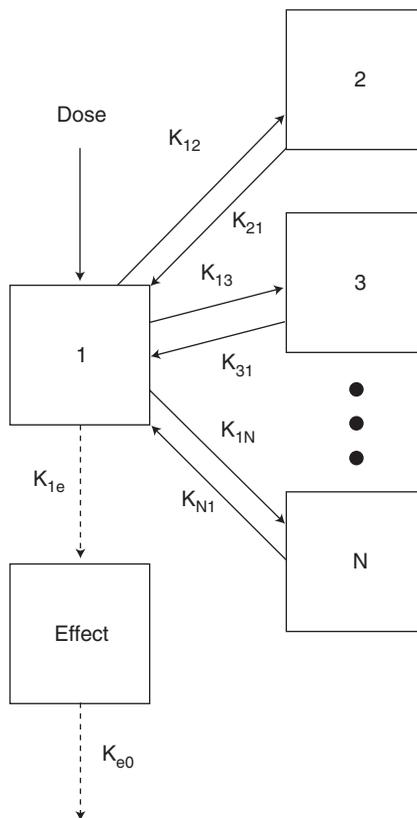
FIGURE 20.7   The PK/PD model developed by Sheiner et al. (6) to couple d-tubocurarine kinetics and dynamics with a link model. The dotted arrow indicates a massless transfer. (From Ref. 6.)

where E is intensity of pharmacologic effect expressed as a fraction of the maximal effect, A_e is effect compartment quantity driving the effect, $A_e(50)$ is analogous to EC_{50}, and γ governs the slope of the log-linear region the described curve. The important PD kinetic parameter in this model is k_{e0} (see Figures 20.6 and 20.7): this parameter alone determines the time delay in this PD model. k_{e0} provides the necessary time delay parameter between introduction of drug and the beginning of the observed effect.

Others have modeled the biophase using systems (24, 25), semiparametric models (50), and nonparametric (15) models to account for the temporal displacement of the effect curve relative to C_p. The two other elements, biosensor and transduction kinetics, in PD modeling can also be employed to account for effect curve temporal displacement. These other two areas are covered in other chapters of this book.

We conclude this chapter with two examples from the literature of highly complex behavior between dose and effect that are modeled with direct PD models. The first example involves the chronopharmacokinetics of the anticancer drug 5-fluorouracil (5-FU). The second example covers nicotine and morphine drug tolerance.

20.4 EXAMPLES OF COMPLEX DYNAMIC BEHAVIOR USING EMPIRICAL/DIRECT MODELS

20.4.1 Circadian Model (Pharmacokinetic Example)

Circadian variation or rhythmicity in humans has been documented for many endogenous substances and physiologic processes (51, 52) as well as for human pharmacokinetics and pharmacodynamics (51, 53). These observations have led to changes in dosing regimens to maximize patient response and to minimize patient toxicity (52, 54). Circadian variation has significant implications in the design of rational drug dosing regimens and therefore in the PK and PD modeling of these circadian-dependent drugs. This section briefly reviews the link between PK and PD models as it relates to time-invariant and time-variant pharmacokinetics/pharmacodynamics. The basic structure of time-variant PK/PD models is then reviewed. Finally, an example of circadian modeling and its circadian–optimized dosing regimen is reviewed.

20.4.1.1 Time-Invariant and Time-Variant Systems
Any PK system that obeys the following criterion is considered time-invariant (TI): for a given PK system, $PK \{dose(t), parameters\} = C_p(t)$; that is, time shifted by an amount τ, the drug plasma concentration $C_p(t)$ must be time shifted exactly by τ, or $PK \{dose(t + \tau), parameters\} = C_p(t + \tau)$. The *parameters* are typically rate coefficients, V_d (apparent volume of distribution), amount or activity of a metabolizing enzyme, and so on and are usually time-invariant (i.e., constant from $t = 0$ to infinity). In time-variant systems (TV), these parameters change as explicit and/or implicit functions of time. Time-variant PK models have the form $PK \{dose(t), parameters(t)\} = C_p(t)$. An even more complicated form of TV models, but probably more mechanistically accurate, has the form $PK \{dose(t), parameters[f(t), t]\} = C_p(t)$, where the *parameters* can vary explicitly or implicitly as a function of time. The following chronopharmacokinetic example has this form, where the clearance varies

with the daily cosine-like changes in enzymatic activity. Although it is sometimes difficult to experimentally determine, time-variant pharmacokinetics can, in fact, be linear or nonlinear with respect to dose. A clearer understanding of these time-variant systems has led to the design of dosage regimens and drug formulations that target high drug concentrations at specific times in the pharmacological biorhythm. The ability to aim for a target concentration at a specific point in time can significantly increase the drug's effectiveness.

20.4.1.2   Time-Variant Pharmacokinetics and Toxicity of 5-Fluorouracil

Several publications have demonstrated circadian variation in the pharmacokinetics and pharmacodynamics of 5-fluorouracil (5-FU) during constant infusions of varying rates typically infused over 5–14 days (55–57). The maximum and minimum concentrations each day based on cosinor analysis occurred at approximately 0100–0400 and 1300 hours, respectively. Dehydropyrimidine dehydrogenase (DPD) is primarily responsible for the metabolism of 5-FU and demonstrates circadian variation in activity with its maximum and minimum activity based on cosinor analysis occurring at 0100 and 1300 hours, respectively. Some patients demonstrated an inverse relationship to the plasma 5-FU concentration (55). This appeared to increase the tolerance to 5-FU side effects between 0000 h and 0400 h (58, 59).

Bressolle and co-workers (60) used a two-cycle, two-amplitude, and two-acrophase model to characterize 5-FU clearance as follows:

$$CL_{ss} = CL_{av} + CL_{A1} \cdot \cos[(t - t_{z1}) \cdot 2/24] + CL_{A2} \cdot \cos[(t - t_{z2}) \cdot 2/12] \qquad (20.17)$$

where CL_{av} is the average clearance; CL_{A1} and CL_{A2} are the amplitudes of the first and second periodic components, respectively; and t_{z1} and t_{z2} are the acrophase (peak) times of the first and the second periodic components, respectively. The addition of the circadian component to CL significantly improved the fit (60). CL varied with sex ($CL_{av} = \text{sex} \cdot \theta_1 + \theta_2$, where sex = 0 if female and 1 if male).

The acrophases or peak times of the 5-FU CL of the first and second cyclic components were at 0414 and 0025 hours, respectively, corresponding to 24 hour minimum concentrations. These times are different from those expected based on the studies discussed previously and reflect the interpatient variability in the circadian rhythm of DPD enzyme activity and 5-FU clearance.

Several investigators have recommended chronomodulating the 5-FU infusion in accordance with the circadian rhythm in 5-FU pharmacokinetics, tumor cell susceptibility, and normal cell tolerability to maximize response and minimize toxicity (54, 57). These authors used a chronomodulated 5-FU infusion from 2200 to 1000 h with a peak infusion rate at 0400 h. 5-FU clinical response was greater and toxicity was less during the chronomodulated infusions versus the continuous infusions.

Characterizing a drug's circadian pharmacokinetics and pharmacodynamics can enable investigators to temporally target the administration time and intensity to maximize patient response and minimize patient toxicity. In these models, the rhythmic displacement in the effect curve is caused by the underlying PK circadian changes in clearance, not any PD interaction. Hence, a direct PD model can still be used to model the PD interaction.

20.4.2 Drug Tolerance (Pharmacodynamic Example)

Drug tolerance is the source of much discussion in the literature. Although many have qualitatively described this phenomenon, until recently, few investigators have attempted to kinetically model or quantitate it, and even fewer have developed rational dosing schemes to circumvent it. For the purposes of this chapter, we define tolerance as "the reversible (relative to the duration of therapy) lessening of drug effect with time, when the drug level is maintained constant." When drug tolerance is present, clockwise hysteresis between drug concentration, C_e, and effect is observed (see Figure 20.8).

The mathematical distinctions between various models provide insight into differentiating between underlying mechanisms of tolerance. For example, some models of tolerance predict identical onset and recovery rates (61–64), while other models predict asymmetrical onset and recovery rates (65–70). For a more complete comparison of tolerance models, the reader is directed to Gardmark et al. (71).

Sheiner and co-workers (61, 62) created a hypothetical kinetic tolerance compartment or state variable connected to a modified E_{max} model (Figure 20.9). The tolerance compartment, denoted *Tol*, is linked to the concentration compartment
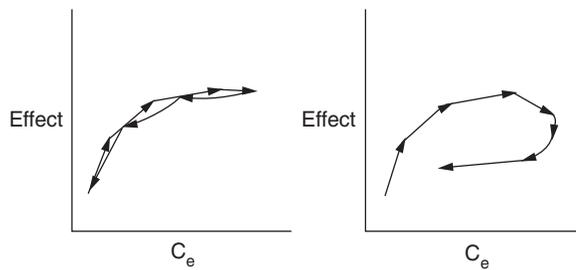
FIGURE 20.8 The left panel shows a typical C_e versus *Effect* plot for a drug obeying a simple E_{max} relationship. The right panel shows the clockwise hysteresis seen in the C_e versus *Effect* plot when drug tolerance is present. The arrows represent the progression of time.
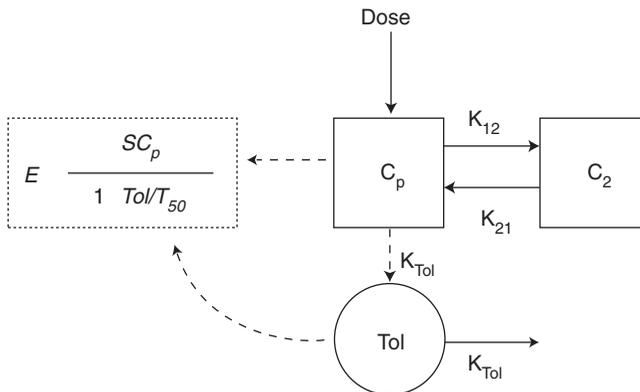
FIGURE 20.9 This tolerance model utilizes a hypothetical compartment (*Tol*) to measure the "amount of tolerance" in the system. *Tol* can act as a competitive or noncompetitive antagonist to the effect. The rate of tolerance development and recovery is determined by k_{Tol} and the effect model is direct. A biophase compartment, C_e, could have also been added (see Figure 20.10). (Adapted from Refs. 61 and 62.)
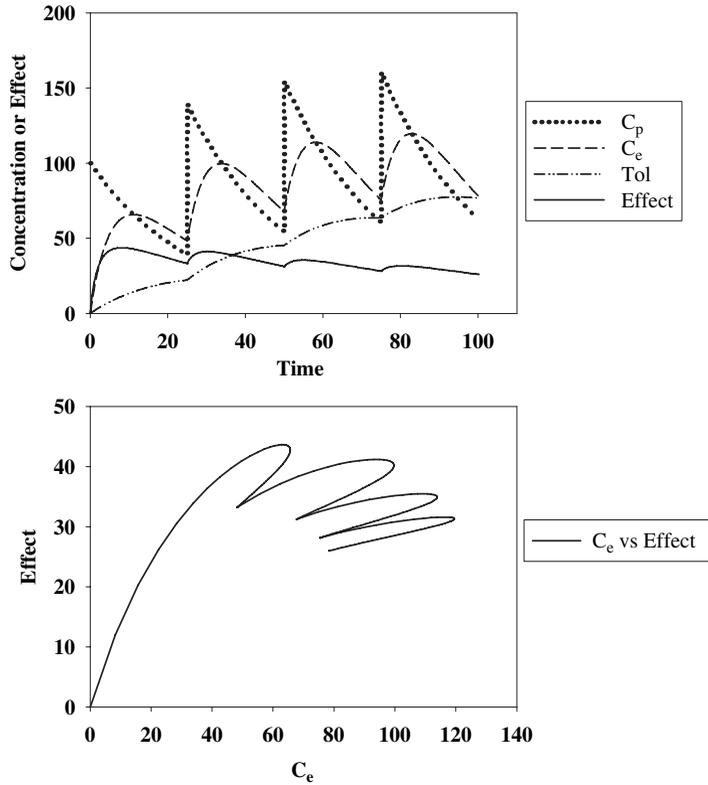
FIGURE 20.10   The upper panel shows four time profiles (C_p, C_e, *Tol*, *Effect*) produced in response to repeated IV bolus doses every 25 time units. The *Tol* curve represents the "amount" of tolerance in the system; as it increases, *Effect* decreases despite increased levels of C_e and C_p. The bottom panel shows a plot of *Effect* versus C_e. Time starts at (0, 0) and moves along the curve. The unusual hand-like shape (hysteresis) is caused by the *Tol* interaction with C_e in the direct PD model. The effect model used was similar to Eq. (20.18), where C_e replaced C_p.

(usually C_p or C) via a first-order transfer constant. The first-order exit rate constant, k_{tol}, for *Tol* determines the rate of tolerance development and recovery. If C_p is the driving compartment, then *Tol* becomes a state variable defined as $Tol = k_{tol}C_p * e^{-k_{tol}t}$ (where $*$ denotes convolution). At any time t', $Tol(t')$ represents a moving, inte-grated exposure history of compartment C_p, which is used to attenuate the system's response at $C_p(t')$. In the simplest model, both C_p and *Tol* are linked together in a direct effect model. The following equations show two PD models—noncompetitive (Eq. (20.18)) and competitive antagonistic inhibition (Eq. (20.19)):

$$E = \frac{E_{\max}\left(\dfrac{T_{50}}{T_{50} + Tol}\right)C_p}{C_{50} + C_p} \tag{20.18}$$

$$E = \frac{E_{\max}C_p}{C_{50}\left(\dfrac{T_{50} + Tol}{T_{50}}\right) + C_p} \tag{20.19}$$

where E_{max} is the maximal effect, C_{50} is the concentration of drug leading to 50% of E_{max} in the absence of tolerance, and T_{50} quantifies the relationship between C_p at steady state and *Tol*. As *Tol* increases, either E_{max} is attenuated (noncompetitive) or C_{50} is increased (competitive); both models predict that *Effect* decreases as *Tol* increases. When *Tol* = 0, both models reduce to the well known E_{max} model. Since k_{tol} solely dictates the kinetics of *Tol*, this model, irrespective of the PD model, predicts that the rate of tolerance development and recovery is symmetrical. This model is a direct effect model because, according to its definition, changes in *Tol* and C_p are instantaneously manifested in *E*.

Interestingly, unlike simple, classic competitive models, increasing drug concentration *C* will not overcome the inhibitor's effect *Tol* because *Tol* will increase proportionately—defeating the gains of increased drug concentration. This model has been used to simulate nicotine tolerance (61), multiple intravenous bolus dose morphine tolerance (64), and tolerance to caffeine's pressor effects (63).

20.5 SUMMARY

Two examples are given to demonstrate that direct effect models can still be used in complex PD behavior. As technological advances in biochemistry and molecular pharmacology techniques continue to provide greater and greater detail of biosensor and transduction systems, these mechanisms need not necessarily be included in an appropriate PK/PD model. When the kinetics of those systems are rapid compared to pharmacokinetics and biophase kinetics, a direct effect PD model may be a better choice—especially when facing the identifiable pitfalls of sparse or incomplete data.

REFERENCES

1. T. Teorell, Kinetics of distribution of substances administered to the body. I. The extravascular modes of administration. *Arch Int Pharmacodyn* **57**:205–225 (1937).

2. T. Teorell, Kinetics of distribution of substances administered to the body. II. The intravascular modes of administration. *Arch Int Pharmacodyn* **57**:226–240 (1937).

3. G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*. Wiley, Hoboken, NJ, 1987, pp. 1–3, 15–33.

4. P. Bonate, W. R. Gillespie, T. Ludden, D. B. Rubin, and D. Stanski, *Simulation in Drug Development: Good Practices*, *Version 1.0*. Draft Publication of the Center for Drug Development Science (CDDS), 1999, pp. 1–19.

5. R. F. Furchgott, The pharmacology of vascular smooth muscle. *Pharmacol Rev* **7**:183–265 (1955).

6. L. B. Sheiner, D. R. Stanski, S. Vozeh, R. D. Miller, and J. Ham, Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to *d*-tubocurarine. *Clin Pharmacol Ther* **25**:358–371 (1979).

7. G. Segre, Kinetics of interaction between drugs and biological systems. *Farmaco (Sci)* **23**:907–918 (1968).

8. W. A. Colburn, Simultaneous pharmacokinetic and pharmacodynamic modeling. *J Pharmacokinet Biopharm* **9**:367–388 (1981).

9. W. A. Colburn, Combined pharmacokinetic/pharmacodynamic (PK/PD) modeling. *J Clin Pharmacol* **28**:769–771 (1988).

10. C. Csajka and D. Verotta, Pharmacokinetic–pharmacodynamic modelling: history and perspectives. *J Pharmacokinet Pharmacodyn* 1–53 (2006).

11. M. Danhof, Applications of pharmacokinetic/pharmacodynamic research in rational drug development. *Methods Find Exp Clin Pharmacol* **18** (Suppl C):53–54 (1996).

12. M. Danhof, J. W. Mandema, A. Hoogerkamp, and R. A. Mathot, Pharmacokinetic–pharmacodynamic modelling in pre-clinical investigations: principles and perspectives. *Eur J Drug Metab Pharmacokinet* **18**:41–47 (1993).

13. H. Derendorf, L. J. Lesko, P. Chaikin, W. A. Colburn, P. Lee, R. Miller, R. Powell, G. Rhodes, D. Stanski, and J. Venitz, Pharmacokinetic/pharmacodynamic modeling in drug research and development. *J Clin Pharmacol* **40**:1399–1418 (2000).

14. N. H. G. Holford and L. B. Sheiner, Understanding the dose–effect relationship: clinical application of pharmacokinetic–pharmacodynamic models. *Clin Pharmacokinet* **6**:429–453 (1981).

15. J. D. Unadkat, F. Bartha, and L. B. Sheiner, Simultaneous modeling of pharmacokinetics and pharmacodynamics with nonparametric kinetic and dynamic models. *Clin Pharmacol Ther* **40**:86–93 (1986).

16. J. G. Wagner, Kinetics of pharmacologic response. I. Proposed relationships between response and drug concentration in the intact animal and man. *J Theor Biol* **20**:173–201 (1968).

17. R. L. Galeazzi, L. Z. Benet, and L. B. Sheiner, Relationship between the pharmacokinetics and pharmacodynamics of procainamide. *Clin Pharmacol Ther* **20**:278–289 (1976).

18. I. S. Nestorov, S. T. Hadjitodorov, I. Petrov, and M. Rowland, Empirical versus mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. *AAPS Pharm Sci* **1**:E17 (1999).

19. W. J. Jusko, H. C. Ko, and W. F. Ebling, Convergence of direct and indirect pharmacodynamic response models. *J Pharmacokinet Biopharm* **23**:5–8; discussion 9–10 (1995).

20. M. Gibaldi and D. Perrier, *Pharmacokinetics*, 2nd ed. *Drugs and the Pharmaceutical Sciences*, Vol. 15. Marcel Dekker, New York, 1982, pp. 45–144.

21. M. Rowland, in *Pharmacokinetics, A Modern View*, L. Z. Benet, G. Levy, and B. L. Ferraiolo (Eds). Plenum Press, New York, 1984, pp. 133–146.

22. M. Rowland and T. N. Tozer, *Clinical Pharmacokinetics: Concepts and Applications*, 3rd ed. Williams & Wilkins, Baltimore, 1995, pp. 137–152.

23. J. W. Mandema, C. D. Heijligers-Feijen, E. Tukker, A. G. De Boer, and M. Danhof, Modeling of the effect site equilibration kinetics and pharmacodynamics of racemic baclofen and its enantiomers using quantitative EEG effect measures. *J Pharmacol Exp Ther* **261**:88–95 (1992).

24. P. Veng-Pedersen, J. W. Mandema, and M. Danhof, Biophase equilibration times. *J Pharm Sci* **80**:881–886 (1991).

25. P. Veng-Pedersen and N. B. Modi, Pharmacodynamic system analysis of the biophase level predictor and the transduction function. *J Pharm Sci* **81**:925–934 (1992).

26. B. E. Dahlstrom, L. K. Paalzow, G. Segre, and A. J. Agren, Relation between morphine pharmacokinetics and analgesia. *J Pharmacokinet Biopharm* **6**:41–53 (1978).

27. F. D. Boudinot, R. D'Ambrosio, and W. J. Jusko, Receptor-mediated pharmacodynamics of prednisolone in the rat. *J Pharmacokinet Biopharm* **14**:469–493 (1986).

28. M. L. Jack, W. A. Colburn, N. M. Spirt, G. Bautz, M. Zanko, W. D. Horst, and R. A. O'Brien, A pharmacokinetic/pharmacodynamic/receptor binding model to predict the

onset and duration of pharmacological activity of the benzodiazepines. *Prog Neuropsychopharmacol Biol Psychiatry* **7**:629–635 (1983).

29. T. P. Kenakin, *Pharmacologic Analysis of Drug–Receptor Interaction*, 3rd ed. Lippincott–Raven Publishers, Philadelphia, 1997, pp. 11–26.

30. D. A. Lauffenburger and J. J. Linderman, *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford University Press, New York, 1993, pp. 135–281.

31. L. D. Limbird, *Cell Surface Receptors: A Short Course on Theory and Methods*, 2nd ed. Kluwer Academic Publishers, Boston, 1996, pp. 27–60.

32. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

33. M. Danhof, G. Alvan, S. G. Dahl, J. Kuhlmann, and G. Paintaud, Mechanism-based pharmacokinetic–pharmacodynamic modeling—a new classification of biomarkers. *Pharm Res* **22**:1432–1437 (2005).

34. D. E. Mager and W. J. Jusko, Pharmacodynamic modeling of time-dependent transduction systems. *Clin Pharmacol Ther* **70**:210–216 (2001).

35. Y. N. Sun and W. J. Jusko, Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* **87**:732–737 (1998).

36. G. Levy, Relationship between elimination rate of drugs and rate of decline of their pharmacologic effects. *J Pharm Sci* **53**:342–343 (1964).

37. G. Levy, Kinetics of pharmacologic effects. *Clin Pharmacol Ther* **7**:362–372 (1966).

38. E. J. Ariëns, *Molecular Pharmacology: The Mode of Action of Biologically Active Compounds*. Academic Press, New York, 1964, pp. 29–42.

39. S. Dutta, Y. Matsumoto, and W. F. Ebling, Is it possible to estimate the parameters of the sigmoid Emax model with truncated data typical of clinical studies? *J Pharm Sci* **85**:232–239 (1996).

40. B. Alexander, D. J. Browse, S. J. Reading, and I. S. Benjamin, A simple and accurate mathematical method for calculation of the EC50. *J Pharmacol Toxicol Methods* **41**:55–58 (1999).

41. F. Patron-Bizet, F. Mentre, M. Genton, C. Thomas-Haimez, and J. Maccario, Assessment of the global two-stage method to EC50 determination. *J Pharmacol Toxicol Methods* **39**:103–108 (1998).

42. P. H. Van der Graaf and R. C. Schoemaker, Analysis of asymmetry of agonist concentration–effect curves. *J Pharmacol Toxicol Methods* **41**:107–115 (1999).

43. L. Aarons, C. Baxter, and S. Gupta, Pharmacodynamics of controlled release verapamil in patients with hypertension: an analysis using spline functions. *Biopharm Drug Dispos* **25**:219–225 (2004).

44. L. Li, M. B. Brown, K. H. Lee, and S. Gupta, Estimation and inference for a spline-enhanced population pharmacokinetic model. *Biometrics* **58**:601–611 (2002).

45. K. Park, D. Verotta, T. F. Blaschke, and L. B. Sheiner, A semiparametric method for describing noisy population pharmacokinetic data. *J Pharmacokinet Biopharm* **25**:615–642 (1997).

46. D. Verotta, Volterra series in pharmacokinetics and pharmacodynamics. *J Pharmacokinet Pharmacodyn* **30**:337–362 (2003).

47. R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling*. Morgan Kaufmann Publishers, Los Altos, CA, 1987, pp. 9–17.

48. R. L. Burden, *Numerical Analysis*, 8th ed. Brooks/Cole, Belmont, CA, 2005, pp. 137–156.

49. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, New York, 1992, pp. 113–116.

50. D. Verotta and L. B. Sheiner, Semiparametric analysis of non-steady-state pharmacodynamic data. *J Pharmacokinet Biopharm* **19**:691–712 (1991).

51. B. Lemmer and B. Bruguerolle, Chronopharmacokinetics. Are they clinically relevant? *Clin Pharmacokinet* **26**:419–427 (1994).

52. F. Levi, Circadian chronotherapy for human cancers. *Lancet Oncol* **2**:307–315 (2001).

53. B. Bruguerolle, Chronopharmacokinetics. Current status. *Clin Pharmacokinet* **35**:83–94 (1998).

54. R. C. Hermida and M. H. Smolensky, Chronotherapy of hypertension. *Curr Opin Nephrol Hypertens* **13**:501–505 (2004).

55. B. E. Harris, R. Song, S. J. Soong, and R. B. Diasio, Relationship between dihydropyrimidine dehydrogenase activity and plasma 5-fluorouracil levels with evidence for circadian variation of enzyme activity and plasma drug levels in cancer patients receiving 5-fluorouracil by protracted continuous infusion. *Cancer Res* **50**:197–201 (1990).

56. G. Metzger, C. Massari, M. C. Etienne, M. Comisso, S. Brienza, Y. Touitou, G. Milano, G. Bastian, J. L. Misset, and F. Levi, Spontaneous or imposed circadian changes in plasma concentrations of 5-fluorouracil coadministered with folinic acid and oxaliplatin: relationship with mucosal toxicity in patients with cancer. *Clin Pharmacol Ther* **56**:190–201 (1994).

57. E. Petit, G. Milano, F. Levi, A. Thyss, F. Bailleul, and M. Schneider, Circadian rhythm-varying plasma concentration of 5-fluorouracil during a five-day continuous venous infusion at a constant rate in cancer patients. *Cancer Res* **48**:1676–1679 (1988).

58. F. Levi, R. Zidani, and J. L. Misset, Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer. International Organization for Cancer Chronotherapy. *Lancet* **350**:681–686 (1997).

59. G. Milano and A. L. Chamorey, Clinical pharmacokinetics of 5-fluorouracil with consideration of chronopharmacokinetics. *Chronobiol Int* **19**:177–189 (2002).

60. F. Bressolle, J. M. Joulia, F. Pinguet, M. Ychou, C. Astre, J. Duffour, and R. Gomeni, Circadian rhythm of 5-fluorouracil population pharmacokinetics in patients with metastatic colorectal cancer. *Cancer Chemother Pharmacol* **44**:295–302 (1999).

61. H. C. Porchet, N. L. Benowitz, and L. B. Sheiner, Pharmacodynamic model of tolerance: application to nicotine. *J Pharmacol Exp Ther* **244**:231–236 (1988).

62. L. B. Sheiner, Clinical pharmacology and the choice between theory and empiricism. *Clin Pharmacol Ther* **46**:605–615 (1989).

63. J. Shi, N. L. Benowitz, C. P. Denaro, and L. B. Sheiner, Pharmacokinetic–pharmacodynamic modeling of caffeine: tolerance to pressor effects. *Clin Pharmacol Ther* **53**:6–14 (1993).

64. D. M.-C. Ouellet and G. M. Pollack, Pharmacodynamics and tolerance development during multiple intravenous bolus morphine administration in rats. *J Pharmacol Exp Ther* **281**:713–720 (1997).

65. H. P. Rang and J. M. Ritter, On the mechanism of desensitization at cholinergic receptors. *Mol Pharmacol* **6**:357–382 (1970).

66. B. Katz and S. Thesleff, A study of the "Desensitization" produced by acetylcholine at motor end-plate. *J Physiol* **138**:63–80 (1957).

67. G. Weiland, B. Georgia, S. Lappi, C. F. Chignell, and P. Taylor, Kinetics of agonist-mediated transitions in state of the cholinergic receptor. *J Biol Chem* **252**:7648–7656 (1977).

68. L. A. Segel, A. Goldbeter, P. N. Devreotes, and B. E. Knox, A mechanism for exact sensory adaptation based on receptor modification. *J Theor Biol* **120**:151–179 (1986).

69. V. Licko, Drugs, receptors and tolerance, in *Pharmacokinetics and Pharmacodynamics of Psychoactive Drugs*, G. Barnett and C. N. Chary (Ed.). Biomedical Publications, Foster City, 1985, pp. 311–322.

70. A. Sharma, W. F. Ebling, and W. J. Jusko, Precursor-dependent indirect pharmacodynamic response model for tolerance and rebound phenomena. *J Pharm Sci* **87**:1577–1584 (1998).

71. M. Gardmark, L. Brynne, M. Hammarlund-Udenaes, and M. O. Karlsson, Interchangeability and predictive performance of empirical tolerance models. *Clin Pharmacokinet* **36**:145–167 (1999).

Developing Models of Disease Progression

DIANE R. MOULD

21.1 INTRODUCTION

21.1.1 Why Do Disease Progression Modeling?

The concept of developing models to describe the progression of disease is not new. Many scientists and clinicians conducted longitudinal studies of the natural time course of disease in the 1970s (1, 2), a trend that has continued although data from untreated patients is difficult to obtain given the wide variety of drugs available to treat most progressive diseases. In the early 1980s Holford and Sheiner (3) suggested a new meaning for the standard E_{max} pharmacodynamic (PD) model, in that the baseline status of a patient, E_0 (or S_0), might not be static, as had been generally assumed in most study designs and analysis of data obtained from clinical trials, and should be taken into account in analysis of such trials. The concept of evaluating drug action on disease trajectory was an important improvement in clinical pharmacology. There are early examples of evaluating drug effect on the natural history of disease (4, 5), although the application of model-based evaluation to disease progression was slow to take hold. Today, however, model-based evaluation of disease progression is an important aspect of drug development and pharmacology. The development of models describing the time course of disease is a component of the critical path initiative described by the US Food and Drug Administration (FDA) in 2004. In short, disease progression modeling has become an accepted tool that should be implemented to evaluate the effects of drug on disease trajectory. Before delving more into this subject, a few questions regarding clinical trials should be addressed.

What is the real purpose of running a clinical trial? In a confirmatory trial, the stated purpose of that trial is usually to test the null hypothesis. Clinical trials are often focused on testing the null hypothesis because there is usually an alternative model, such as that the drug does have clinically relevant activity, that can be accepted in place of the null model. Furthermore, testing the null hypothesis is an

easy question to answer robustly, and traditionally statistics has been focused on questions that are easy to answer but not necessarily on answering the right questions. However, as Tukey (6) reminded us, "far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise." Therefore, the consequences of addressing the more interesting general questions are that the answers are generally less robust. For some time now, Sheiner (7, 8) and others (9) have argued that greater statistical power can be achieved by addressing these more interesting questions through the use of alternative hypotheses that test accepted scientific models of disease and response rather than limiting the questions to testing the null hypothesis.

Ideally then, generating a summary of evidence of efficacy is more useful, although that implies developing an understanding about the dose–response curve in the target patient population. If the objective of addressing the more general question, such as understanding the dose response, is accepted, then a dose–response surface must be developed. This response surface is a multivariate response surface that provides the expected value of response as a function of multiple covariates and dosing levels and is described by Sheiner (10) as a part of the learning and confirming process in drug development.

Developing the response surface is a difficult question to address and usually requires that a variety of assumptions be made when addressing it, which weakens the robustness of the answers. However, a summary of the surface function, such as an average over that response surface function, often can provide robust answers to simpler questions. The margins of a high-dimensional surface are often very well estimated and robust, even when using a model to describe the data. Therefore, if a model is used to address the right question, the answer will have uncertainty associated with that answer, but summarizing or integrating over that model in order to answer simpler questions can still provide robust answers.

The consequences of providing less robust answers usually will not adversely impact labeling or clinical practice. In the context of answering questions such as how the drug works in a specific subpopulation of patients or how to adjust the starting dose, not providing information forces the clinicians to make decisions based on no information at all. Therefore, it can be argued that providing somewhat less robust answers also provides a somewhat better solution to the problem of providing information about how to use the drug in practice.

The reason for addressing these more complex questions is because during the development of a new drug candidate, different patients often have different responses to the same dose of a test drug. For example, older patients may be more or less sensitive to both the positive and negative effects of drugs. This difference in sensitivity to a test drug contributes to the variability (e.g., noise) in the outcome of the study. Unfortunately, however, it is usually not possible to study all combinations of doses or treatments by patient type to determine explicitly what the dose response is for each subgroup of patients that may be administered the drug. Therefore, it may be necessary to describe the dose–response surface without data from every type of patient given every dose level and duration of therapy. Model-based evaluations, however, can provide a basis for developing a dose–response surface by making scientifically valid assumptions, without which, model-based analysis cannot proceed. These models can then be used for interpolation and, in some cases, for extrapolations as well.

Assumptions are included in all of the elements of any pharmacokinetic/pharmacodynamic (PK/PD) model. Some examples of common assumptions made for these models include the structure of the models for pharmacokinetics, pharmacodynamics, and their respective covariate influences, the models for the clinical effect of the drug, the parameter values of all these models, and the variance structures for model components (11). Assumptions reduce inferential certainty because if the assumptions are wrong, then the model-based conclusions are wrong. Therefore, it is the quality of the attendant assumptions, and not their existence, that is the issue with assumptions in modeling (12).

In addition to providing a basis for interpolation and extrapolation, models increase the amount of information recovered from a clinical trial. Information obtained from any scientific study can be detected based on the ratio of signal to noise. In any given study, the information is the total variation in the data, the signal is the variation due to identifiable causes such as differences in dose, and the noise is the residual or unexplained variation. Therefore, models increase information by turning noise into signal by providing a basis for explaining the variation.

Several models are necessary to generate the response surface. These include the pharmacokinetic (PK) model (what the body does to the drug over time), the pharmacodynamic (PD) model (what the drug does to the body over time), and a disease progression model (how the disease is changing over time). A disease progression model is actually a special class of PD models in that it is a function that describes the change of the clinically relevant endpoint over time. The response surface then is a function of all these three models. Therefore, development of a disease progression model is a valuable tool to facilitate the visualization of underlying disease changes for both the reference treatment and the treatment being evaluated. Disease progression models are an important component of the drug exposure–response surface.

21.1.2 Background

Making relevant scientific assumptions involves understanding the mechanistic relationship between drug treatment and observed responses. In the context of modeling the response surface, it is useful to think of clinical pharmacology as the combination of disease progress and drug action (13).

$$\text{Clinical pharmacology} = \text{Disease progress} + \text{Drug action} \qquad (21.1)$$

In this framework, disease progress refers to the evolution of a disease over time, or the disease trajectory, which can be assessed by observing the time course of a biomarker or other clinically relevant endpoint that reflects the status of a disease or is a measure of the clinical status of a patient. The status of the patient is a reflection of the state of the disease at a point in time. Disease status may improve or worsen over time, or may be a cyclical phenomenon such as the seasonal affective disorder component of depression. Therefore, a model of disease progress is a mathematical expression that describes the expected changes in patient status over time either in the absence of treatment or at least in the absence of the treatment being investigated.

Disease progress models can be extended to include terms that account for the changes in disease progress that are impacted by drug treatment or drug action. Drug action refers to all the underlying PK and PD processes involved in producing a drug effect on the disease progression as these are the two major attributes determining drug exposure and its effect on the time course of progression of the disease.

21.2   DATA

21.2.1   Pooling Data

Because disease progression is usually measured by observational data such as scores or summary measures, the data are usually quite variable. Furthermore, the progression of many diseases is often quite slow, requiring observational data collected over long periods of time. As a consequence, it is often necessary to pool data from several clinical studies of the disease in order to develop a model for disease progression.

There are potential problems associated with pooling data from multiple studies: different treatments and duration of treatments, assessment of patients using different schedules, and different enrollment criteria for the protocols. Using models to analyze data helps to adjust for design problems, however. Explicit modeling of covariate effects, such as sex, age, or other important patient characteristics, allows the analyst to pool data from different patient groups and different trials because the model can estimate how to adjust for the different groups. Using time as a covariate in the model system facilitates pooling data from studies with different treatments and different designs. Explicit modeling of variation in error structures allows pooling data with different precision, such as might arise from different assays.

During any clinical trial, several clinical endpoints and measurements of relevant biomarkers are collected for all patients, regardless of treatment assignment. Such observations can be either subjective (such as assessments of how a patient feels or functions) or objective (such as the evaluation of biomarkers). The selection of the biomarker or endpoint that is to be modeled depends on how well the overall disease progression is represented by these data. The selection may also take into account the amount of data available for each observation type and to some extent the objectives of the modeling exercise. It is important to also include data from the untreated (placebo) patients in order to evaluate the time course of disease in the absence of test drug. In many cases, the placebo patients will have received treatment that is the standard of care for their disease (also called the active control arm). This standard therapy is usually also administered to the patients receiving the test medication. If the active control is not also administered to the patients receiving the test medication, then combining data from such treatments might not be appropriate. Whenever data from multiple studies are pooled together, careful consideration needs to be given as to whether or not the differences in design and treatment can be described sufficiently to allow them to be combined.

TABLE 21.1   Example Database Format for Disease Progression Model

Patient ID	Date	Time	AMT	DV	DVID	MDV	EVID
1	7/31/2000	13:20	3,000	•	0	1	1
1	7/31/2000	13:20	•	124	1	•	•
1	8/2/2000	10:15	30,000	•	0	1	1
1	8/2/2000	12:47	•	700	1	•	•
1	8/7/2000	10:45	30,000	•	0	1	1
1	8/7/2000	10:45	•	37	2	•	•
1	8/7/2000	13:15	•	729	1	•	•
2	1/3/1992	8:00	•	6	2	•	•
2	2/3/1992	15:00	•	8	2	•	•
2	2/3/1992	15:15	7,500	•	0	1	1
2	2/6/1992	16:59	•	7	2	•	•
2	2/10/1992	8:50	7,500	•	0	1	1
2	2/10/1992	10:05	•	5	2	•	•
2	2/13/1992	0:00	•	6	2	•	•
2	2/17/1992	8:36	•	6	2	•	•
3	1/9/1992	11:06	•	117	1	•	•
3	1/27/1992	8:21	•	176	1	•	•
3	1/27/1992	11:30	7,500	•	0	1	1
3	1/27/1992	11:43	•	155	1	•	•
3	1/27/1992	14:30	•	81	1	•	•
3	1/27/1992	16:30	•	91	1	•	•

21.2.2   Database Formatting

In many disease progression models, integrated functions are used to describe both the pharmacokinetics and pharmacodynamics of the disease progression. For this example database format, it is assumed that the data will be fit using integrated equations. An example of a standard format for a disease progression model is provided in Table 21.1. In this example, both the PK observations and the markers for disease progression are listed in the DV column. A flag (DVID) indicating whether the record is a dose record (DVID = 0), a concentration record (DVID = 1), or a disease progression record (DVID = 2) is necessary both for application of the appropriate residual error function and for sorting the output and examining model performance afterwards. Additional columns for covariates can be added as needed.

21.3   MODELS

There are several models that need to be developed to describe the exposure–response surface. These include both PK and PD models, where the PD model includes models of disease progression and drug action. Prior to developing a model for disease progression, it is helpful to examine the different components of a disease progression model, and to understand the terminology associated with these models.

21.3.1 Definitions

The terminology presented here is consistent with terminology described and used previously (13, 14). In order to demonstrate the different types of drug action, example plots were generated using a simple linear disease progression model.

1. *Symptomatic.* A drug action is defined as being "symptomatic" when it has a beneficial effect on the disease status but does not alter the trajectory the disease. An example of a drug action that has symptomatic benefit is shown in Figure 21.1: during the period of time that drug concentrations are non zero, the disease status is lower in the treated arm than in the untreated arm. However, the slope of the line describing the treated arm is the same as the untreated arm, suggesting that the disease is progressing. As further proof that the drug is not altering the progression of disease, when the drug treatment is stopped, the status of treated arm rapidly becomes indistinguishable from the untreated arm. Therefore, the presence of the drug ameliorates the disease but does not alter its time course.

2. *Disease Modifying.* A drug is said to exhibit disease modifying action (referred to as having "protective action") if the drug alters the progression of disease. Consequently, the effects of the drug action persist even after the drug is removed. There are two general types of protective actions, the first being one that alters the time course of the disease while the second action involves altering the maximal status of the disease. An example of a protective agent that alters the time course of disease is shown in Figure 21.2.

3. *Curative.* A curative drug completely halts the progression of a disease and reverses the patient status back to the predisease state. Even after cessation of therapy, the patient status remains at the predisease state. An example of a curative drug is provided in Figure 21.3. As can be seen in this figure, the baseline disease status reverses rapidly to 0 and remains there after removal of drug.
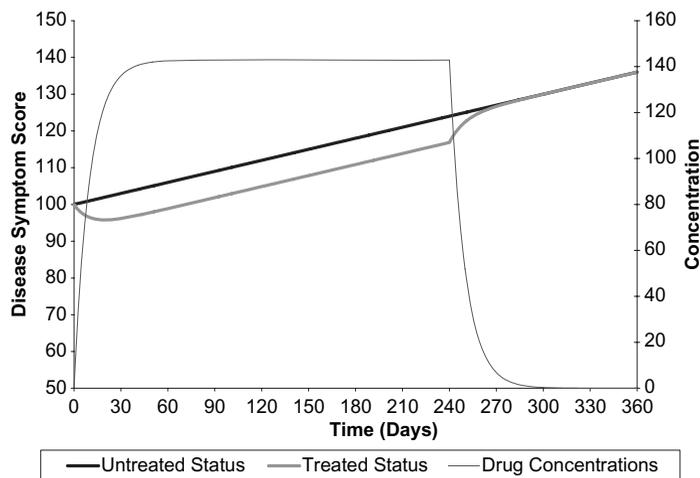
FIGURE 21.1 Example profile of symptomatic drug action.
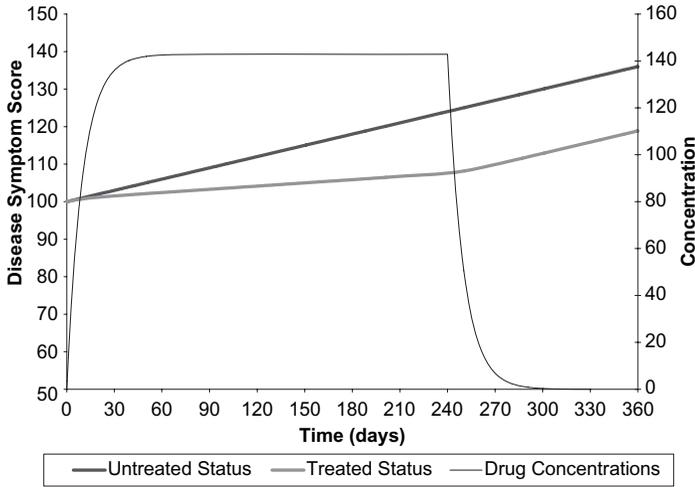
FIGURE 21.2   Example profile of disease progression modifying drug action.
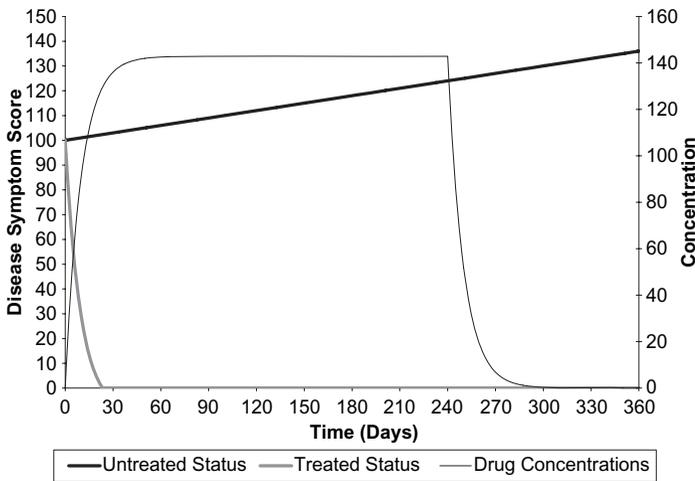
FIGURE 21.3   Example profile of curative drug action.

21.3.2   Structural Models of Disease Progression

21.3.2.1   Linear Models of Disease Progression

The linear disease progression model is the simplest model that is used to describe disease progression. Although the linear model of disease progression is quite simplistic, it has been used to describe the progression of several different diseases. The linear model for disease progression was used by Holford and Peace to describe the progression of Alzheimer's disease (15). This model was developed to allow evaluation of all patients treated with the drug and was used to determine the effectiveness of tacrine as a treatment for patients with the disease.

The linear model assumes a constant rate of change of the disease status over time. The linear model can be defined in terms of a baseline disease status (S_0) and a slope parameter (α) and t is time after the initial observation of the disease. The equation for the linear model is given below and a NONMEM control stream implementing the linear disease progression model is provided in Table 21.2.

$$S(t) = S_0 + \alpha t \tag{21.2}$$

When building a model for disease progression, it is often best to develop the disease progression model first, and then a model for drug effect is added. Typically, several disease progression models will be tested and the one that appears to best describe the time course for the markers of patient status is taken further to evaluate the addition of models for drug effect.

With a linear model, there are two basic drug effect patterns possible. Symptomatic drug action will improve patient status but has no impact on the rate of progress, or the drug can alter the rate of progress of the disease, resulting in a protective or disease modifying action.

A symptomatic benefit can easily be described by adding an effect based on drug concentration. In this case, the drug effect, $E(t)$, modifies the patient status by shifting it by a constant amount over time as long as drug is present.

$$S(t) = S_0 + E(t) + \alpha t \tag{21.3}$$

The expected pattern of drug action then would be expected to be similar to the example provided in Figure 21.1. For the situation where the drug effect appears to modify the progression of the disease, the drug effect, $E(t)$, can act on the slope parameter. This effect would be expected to be similar to the example provided in Figure 21.2.

$$S(t) = S_0 + [E(t) + \alpha]t \tag{21.4}$$

The symptomatic and disease modifying drug effects can also be combined if there is evidence that the drug exhibits both types of activity. Again, good model building practices should be employed to test if the combined symptomatic and modifying drug effect model is more appropriate than either effect alone.

The drug effect, $E(t)$, can be described using a linear function, an E_{max} function, or other function, as is appropriate.

The onset of drug effect may be delayed by adding a hypothetical effect compartment to the drug action part of the model, and using the concentrations at the effect site to be the forcing function for the drug effect (17, 18). In general, a delay in the onset of drug effect is expected although the appropriateness of this assumption should be tested using good model building practices. As an additional caution, however, when using models to assist in determining the mechanism of action of a drug, it should be noted that a drug that has symptomatic activity but a long delay to the onset of effect can provide a response that is indistinguishable from disease modifying activity. Figure 21.4 shows a linear disease progression model with treated and untreated status. In this simulated scenario, the pharmacokinetics of the drug used to drive the response are identical, however, the lag time to onset

TABLE 21.2 Example NONMEM Code 1: Linear Model of Disease Progression with No Drug Effect

NONMEM Code	Explanation
| ```
$SUBROUTINE ADVAN2 TRANS2
$PK
  CALLFL=-2
  IF (NEWIND.LE.1) THEN
  LN2=LOG(2)
  TWOPI=2*3.141592654
  DOSE=0
  ENDIF
``` | In this simple example, the drug effect has not yet been added but the PK model is included in the control stream. The model calls for a linear one-compartment model with first-order input. CALLFL=-2 is a useful call to make particularly if CALLFL controls when PREDPP calls the PK and ERROR routines. CALLFL=-2 will call the PK subroutine with every event record, and with additional and lagged doses. NEWIND is an indicator variable that can be used to calculate values at the first record in the database (NEWIND = 0) or at start of a subsequent individual (NEWIND=1). |
| ```
;normalize covariates
NWT=(WT/70)
IF (WT .LE. 0) NWT=1
NAGE=AGE/50
IF (AGE .LE. 0) NAGE=1
``` | Note that disease progression model control streams can become quite long. Use comment lines to remind yourself or assist others in determining what is going on. Normalizing covariates is always a good idea. This benefits the user in two ways: first by providing some numerical stability and second by allowing the parameter estimates to reflect the subject demographics that reflect the "average" patient. |
| ```
;PK MODEL
TVCL=THETA(1)
CL=TVCL*EXP(ETA(1))
TVV=THETA(2)
V=TVV*EXP(ETA(2))
TVKA=THETA(3)
KA=TVKA*EXP(ETA(3))
S2=V2
``` | Again, commenting within the control stream is a good idea. Note that the parameters associated with the PK model have been fixed. This is consistent with the approach suggested by Zhang et al. (16) for pharmacokinetic and pharmacodynamic evaluation. |
| ```
$ERROR
  CP=A(2)/S2
;BEGIN DISEASE PROGRESS
  MODEL
  TVS0=THETA(4)
  S0=TVS0*EXP(ETA(4))
  TVSLO=THETA(5)
  SLOPE=TVSLO*EXP(ETA(5))
``` | Because the disease progression is not part of the PK model, it must be defined in the $ERROR routine. For simplicity, concentration (CP) is defined here as the ratio of A(2)/S2; then the disease progression parameters are defined. TVS0 is the typical value of the baseline status, S0 is the individual baseline status, TVSLO is the typical value of the slope of progression, and SLOPE is the individual progression. |
| ```
;Derived values
  TDAY=TIME/24 ; convert
    hours to days
  REC=SLOPE*TDAY
``` | TDAY is used here instead of TIME, which is usually in hours for the PK model. Using TDAY gives the disease progression model in days (or any unit of time such as weeks or months) rather than hours for several reasons. The first reason is that the parameter estimates are usually more |

TABLE 21.2 *Continued*

NONMEM Code	Explanation
	meaningful in days rather than in hours. The second reason is that the larger slope estimate for days is generally easier to estimate (and more numerically stable) than the value of the slope in hours (which would be quite small).

``IF (DVID .LE. 1) THEN
  Y=CP*EXP(CVCP) + SDCP
ELSE
;DISEASE SCORE VS TIME
  (days)
  Y=S0-REC + SD
ENDIF```
	DVID was a data flag used to differentiate data records for dosing (DVID=0), for concentrations (DVID=1), or for disease score (DVID=2). The use of the flag is important for separating the observations for residual error and also for plotting and other diagnostic purposes.

``;PARAMS FOR PK
$THETA 20 FIX; CL
$THETA 70 FIX; V2
$THETA 0.1 FIX; KA
;PARAMS FOR PD MODEL
$THETA (10,50,100);
  POPS0 days
$THETA (0,1,100);
  popslo u```
	Note again here that the parameters are fixed for the PK model. For the PD (disease progress) model the initial estimate for the baseline status (S0) is the average value at study entry in the data. The initial estimate of slope can be obtained by plotting the raw data and conducting a linear regression to get a slope estimate.

```;VARIANCE FOR PK $OMEGA .1 FIX; BSVCL $OMEGA .1 fix; BSVV2 $OMEGA .1 FIX; BSVKA ;VARIANCE FOR PD $OMEGA 1;PPVS0 $OMEGA 1;PPVslo```	The variance terms for the PK model were fixed. The variance terms for the PD model are usually larger than for the PK model. Reasonable estimates are usually 0.5 or sometimes larger.
```;RESIDUAL FOR PK $SIGMA 0.15 FIX; CVCP $SIGMA 1 FIX; SDCP ;RESIDUAL FOR PD $SIGMA 10;SD```	
```$EST MAX=9990 SIG=3 NOABORT PRINT=1 METHOD=COND MSFO=base.msf```	The conditional estimation method is used here because the residual error for the disease progression model is additive. The conditional method with interaction can also be used as well. Because disease progression models can run for extended periods of time due to complex models and the large databases required, the use of the MSF file option is recommended. This option allows the job to be restarted if the minimization process is terminated for some reason (e.g., power failures).

FIGURE 21.4   Long delay to onset of effect versus short delay to onset.

is different. In the drug with a short lag time, the slope of progression is clearly following the untreated progression and removal of drug returns the status to the untreated state. However, in the case of the long lag time, the slope of progression appears different from that of the untreated state and removal of treatment does not result in a rapid return to the untreated status. In the second case with a long lag time, it seems likely that a drug effect model with disease modifying activity would describe these data. Therefore, the ability to distinguish whether or not a drug exhibits disease modifying activity is largely dependent on the length of time that the disease progression is monitored.

An example fragment of a NONMEM control stream for a linear model of disease progression with a symptomatic drug effect is provided in Table 21.3, and an example is included in the appendix.

Another example of the use of the linear disease progression model is the work of Kimko et al. (19), describing the recovery following an acute psychiatric episode of schizophrenia using the Brief Psychiatric Rating Score (BPRS). The BPRS is one of the most widely used rating scales in psychiatry. The scale is comprised of 16 items rated from 0 (not present) to 6 (extremely severe). Interpretation of the total scores is: 0–9, not a schizoaffective case; 10–20, possible schizoaffective case; 21 or more, definite schizoaffective case. Consequently, the BPRS score cannot go lower than 0. The application of a different function to describe recovery might have seemed more appropriate but was justified by the model performance, which agreed well with the Phase 3 results from a different study.

In some cases, therefore, the linear model can be a reasonable approximation for disease progression when the disease is observed over a relatively short period of time or if the number of individuals experiencing complete recovery is limited. For example, the trajectory of the United Parkinson's Disease Rating Score (UPDRS— a measure of Parkinson's disease) has been described using Gompertz functions (20), which exhibit an asymptotic increase to a maximum score. However, over

TABLE 21.3   Example NONMEM Code 2: Linear Model of Disease Progression with Symptomatic Drug Effect

Nonmem Code	Explanation

```$SUBROUTINE ADVAN4```
```$PK```
``` CALLFL=-2```
``` IF (NEWIND.LE.1) THEN```
``` D2H=24```
``` LN2=LOG(2)```
``` TWOPI=2*3.141592654```
```ENDIF```
	Now the drug effect has been added using the effect compartment. The PK model is still a linear one-compartment model with first-order input. However, a two-compartment model is being used to allow the second (peripheral) compartment to be the hypothetical effect compartment. Note that in this case, the default TRANS is 1, and that is being used because we require an estimate of the inter compartmental transfer rate constants as keo.

```;normalize covariates```
```NWT=(WT/70)```
```IF (WT .LE. 0) NWT=1```
```NAGE=AGE/50```
```IF (AGE .LE. 0) NAGE=1```	

```;PK MODEL```
```TVCL=THETA(1)```
```CL=TVCL*EXP(ETA(1))```
```TVV=THETA(2)```
```V=TVV*EXP(ETA(2))```
```TVKA=THETA(3)```
```KA=TVKA*EXP(ETA(3))```
```TVTEQ=THETA(4)```
```TEQ=TVTEQ*EXP(ETA(4))```
``` K=CL/THETA(2)```
``` K23=.001*K ;SET TO LOW```
```VALUE SO THAT LOSS TO```
```EFFECT COMT IS TRIVIAL```
``` K32=LN2/TEQ```
``` S2=THETA(2)```
``` S3=S2*K23/K32```
	Note again that this control stream is becoming longer. Commenting becomes essential.
	Once again the parameters associated with the PK model have been fixed. Note that a parameter for the equilibrium half-life (TEQ) has been defined. The value for keo can also be defined.
	Because of the need to use the two-compartment model with TRANS1 for the PK in order to use an effect compartment , microconstants have to be determined. In order to keep the loss of drug to the effect compartment trivial, the transfer rate constant is fixed to a small value. The return value K32 is set to LN2/TEQ and the scale parameter for the effect compartment is defined.

```$ERROR```
``` CP=A(2)/S2```
``` CE=A(3)/S3```
```;BEGIN BPRS MODEL```
``` TVS0=THETA(5)```
``` S0=TVS0*EXP(ETA(5))```
``` TVSLO=THETA(6)```
``` SLOPE=TVSLO*EXP(ETA(6))```
	Concentration (CP) and effect site concentration (CE) are defined here as the ratio of A(2)/S2 and the ratio of A(3)/S3.
	Then the disease progression parameters are defined as before.

```;Derived values```
``` TDAY=TIME/24 ; convert```
```   hours to days```
``` REC=SLOPE*TDAY```
	Change the time scale again to days and calculate the recovery function.

```;DRUG EFFECT```
``` TVDMAX=THETA(7)```
``` DMAX=TVDMAX*EXP(ETA(7))```
``` TVDC=THETA(8)```
``` DC50=TVDC*EXP(ETA(8))```
``` DOFF=DMAX*CE/(DC50+CE)```
	Now we define the drug effect parameters. Here the drug effect will be described using an E_{max} model. DOFF is the effect of drug.

TABLE 21.3   *Continued*

NONMEM Code	Explanation
| ```
IF (DVID .LE. 1) THEN
  Y=CP*EXP(CVCP) + SDCP
ELSE
;DISEASE SCORE VS TIME
  (days)
  Y=S0-REC+DOFF + SD
ENDIF
``` | The appropriate residual error is determined as before with the DVID flag. Note that now the drug effect DOFF is being added to the disease score model, is a SYMPTOMATIC model for drug activity. |
| ```
;PARAMS FOR PK
$THETA 20 FIX; CL
$THETA 70 FIX ; V2
$THETA 0.1 FIX ; KA
;PARAMS FOR LAG
$THETA (0,24,) ;TEQ
;PARAMS FOR PD MODEL
$THETA (10,50,100) ;
  POPS0 days
$THETA (0,1,100) ;
  popslo u
;PARAMS FOR DRUG EFFECT
$THETA (-INF,-.1,0)
  ;DMAX
$THETA (0,2,) ;DC50
``` | Note that the TEQ has been added as have the parameters for drug effect Drug effect DMAX is allowed to go negative here because for this example, a lower score is better. We assume that the drug effect is to lower the score. |
| ```
;VARIANCE FOR PK
$OMEGA .1 FIX ; BSVCL
$OMEGA .1 fix ; BSVV2
$OMEGA .1 FIX ; BSVKA
;VARIANCE FOR LAG
$OMEGA ,1 ;BSVTLAG
;VARIANCE FOR PD
$OMEGA 1 ;PPVS0
$OMEGA 1 ;PPVslo
;VARIANCE FOR DRUG
  EFFECT
$OMEGA 1 ;PPVDMAX
$OMEGA 1 ;PPVDC50

;RESIDUAL FOR PK
$SIGMA 0.15 FIX ; CVCP
$SIGMA 1 FIX ; SDCP
;RESIDUAL FOR PD
$SIGMA 10 ;SD

$EST MAX=9990 SIG=3
  NOABORT
PRINT=1 METHOD=COND
MSFO=base.msf
``` | The variance terms for the PK model were fixed as before. The variance terms for the PD and drug effect models are usually larger than for the PK. Reasonable estimates for all these are usually 0.5 or sometimes larger. Usually the PD variance term estimates can be updated based on the initial runs of the model without drug effect added in. |

short intervals of time, the UPDRS can be approximated using simpler functions. This approximation has the obvious limitation of not being appropriate for long-range predictions, however. Whenever a disease progression model is developed, the limitations of the model need to be evaluated.

21.3.2.2 Asymptotic Models of Disease Progression
The asymptotic models are so named because the models contain an inherent maximal or minimal value that the function slowly approaches as time increases (e.g., the asymptote). There have been several such models proposed for various disease states where there is a natural limit in the progression of disease.

Exponential Function   In some cases of disease progression, such as recovery from an injury or some other temporary disease state, the model should be able to describe the improvement over time. In such cases, recovery can be approximated by an exponential function parameterized for the baseline status S_0 and the rate constant of recovery k_{prog}. The exponential function has the property of asymptotically approaching 0 and so is best used in situations where the severity scores have a minimum value of 0 or, in the case of some biomarkers, do not occur in the nondiseased state.

$$S(t) = S_0 e^{-k_{prog}t} \tag{21.5}$$

With the exponential asymptotic model, the effect of drug can be described as symptomatic or as disease modifying or as a combination of both. In the case of the symptomatic benefit, as was seen with the linear model for progression, the drug effect, $E(t)$, is added directly to the function for disease status.

$$S(t) = S_0 e^{-k_{prog}t} - E(t) \tag{21.6}$$

A drug that exhibited disease modifying behavior would be expected to impact the rate constant of recovery by increasing that rate constant, thereby shortening recovery time.

$$S(t) = S_0 e^{-(k_{prog}t + E(t))t} \tag{21.7}$$

An example profile for an asymptotic recovery model with and without a symptomatic drug effect model is presented in Figure 21.5, and the same asymptotic model with and without disease modifying drug effect is presented in Figure 21.6. It is worth noting that the recovery time for the untreated case is much shorter than for other diseases. Consequently, a drug with disease modification characteristics may offer less of a clinical advantage than one that offers rapid symptomatic benefit.

A portion of the NONMEM code that will describe an asymptotic exponential disease progression model with a symptomatic drug effect model is provided in Table 21.4. A portion of the code for the same disease progression model with the drug effect described as disease modifying is provided in Table 21.5.

Several different models have been developed for applications in osteoporosis. An exponential model with and without linear components was evaluated by Pors Neilsen et al. (21) to describe changes in bone mineral density, while Pillai et al.

FIGURE 21.5   Example profile of asymptotic exponential model with symptomatic drug action.

FIGURE 21.6   Example profile of asymptotic exponential model with disease progression modifying drug action.

(22) investigated the use of an indirect response type model. However, it should be noted that the markers of disease progression in these two examples were quite different. In the first, bone mineral density of the lumbar spine was evaluated over time and in the second example, the model was developed to describe the urinary excretion of the C-telopeptide of the A chain of type I collagen. Consequently, the selection of an appropriate model needs to be based on the marker of disease progression being described.

TABLE 21.4   Example NONMEM Code 3: Asymptotic Exponential Model of Disease Progression with Symptomatic Drug Effect

NONMEM Code	Explanation

<pre>$ERROR
  CP=A(2)/S2
  CE=A(3)/S3
;BEGIN DISEASE PROGRESSION
MODEL
  TVS0=THETA(5)
  S0=TVS0*EXP(ETA(5))
  TVKPROG=THETA(6)
  KPROG=TVKPROG*EXP(ETA(6))
  REC=EXP(-KPROG*TIME)
;DRUG EFFECT
  TVDMAX=THETA(7)
  DMAX=TVDMAX*EXP(ETA(7))
  TVDC=THETA(8)
  DC50=TVDC*EXP(ETA(8))
  DOFF=DMAX*CE/(DC50+CE)</pre>
	In this example, we are again testing a delay to drug effect. The approach is the same as was shown previously, to use an ADVAN that includes an extra compartment for the effect site. Again, because the model involves the use of an ADVAN (integrated function), the disease progression model must be defined in the $ERROR routine. In this case REC, the recovery parameter, is defined as an exponential function.
	
	 Again, we define the drug effect parameters. Here the drug effect will be described using an E_{max} model where DOFF is the effect of the drug.

<pre>IF (DVID .LE. 1) THEN
  Y=CP*EXP(CVCP) + SDCP
ELSE
;DISEASE SCORE VS TIME (days)
  Y=S0*REC+DOFF + SD
ENDIF</pre>
	The appropriate residual error is determined as before with the DVID flag. Note that now the drug effect DOFF is being added to the disease score model. This describes *symptomatic* drug activity.

TABLE 21.5   Example NONMEM Code 4: Asymptotic Exponential Model of Disease Progression with Disease Modifying Drug Effect

NONMEM Code	Explanation

<pre>$ERROR
  CP=A(2)/S2
  CE=A(3)/S3
;DRUG EFFECT
  TVDMAX=THETA(5)
  DMAX=TVDMAX*EXP(ETA(5))
  TVDC=THETA(6)
  DC50=TVDC*EXP(ETA(6))
  DOFF=DMAX*CE/(DC50+CE)</pre>
	We define the drug effect parameters. Here the drug effect will be described using an E_{max} model where DOFF is the effect of drug. This describes DISEASE MODIFYING drug activity

<pre>;BEGIN DISEASE PROGRESSION
MODEL
  TVS0=THETA(7)
  S0=TVS0*EXP(ETA(7))
  TVKPROG=THETA(8)*(1+DOFF)
  KPROG=TVKPROG*EXP(ETA(8))
  REC=EXP(-KPROG*TIME)</pre>
	In this case REC, the recovery parameter, is defined as an exponential function. However because the model that is being tested involves a disease modifying activity, the effect needs to be added to the TVKPROG.

<pre>IF (DVID .LE. 1) THEN
  Y=CP*EXP(CVCP) + SDCP
ELSE
;DISEASE SCORE VS TIME (days)
  Y=S0*REC + SD
ENDIF</pre>	

E_{max} Functions   Another familiar asymptotic function is the "E_{max}" function, which has a natural limit S_{max}. The E_{max} model for disease progression has been used to describe the progression of several different disease scores that have a natural limit associated with the score. Anderson et al. (23) used the E_{max} model to describe pain resolution in pediatric patients and Taylor et al. (24) used this model to describe recovery from ischemic stroke using the National Institutes of Health Stroke Score. This model adequately described the trajectory of both markers of disease progression and was able to describe wide interpatient variability in disease progression and response.

The E_{max} model is a simple function to implement in NONMEM and has the advantage of parameterization that is reasonably familiar, making the modeling results relatively easy to interpret by individuals who are not familiar with modeling. The patient status at any time "t" is described as the sum of the baseline status S_0 and some recovery function that has a maximum of S_{max}. The time to half maximal recovery is S_{50}.

$$S(t) = S_0 + \frac{S_{max}t}{S_{50} + t} \tag{21.8}$$

For situations where the change in disease status over time is rapid, a Hill coefficient can be included. In addition to the E_{max} function, which has an asymptotic increase, an "I_{max}" function can also be used, which describes an asymptotic reduction in disease score over time. Like the exponential function, the I_{max} function will asymptotically approach a value of 0.

As with all the previous functions, drug effect can be included as symptomatic, disease modifying, or a combination of both. Symptomatic drug action is simply added to the function. For disease modifying agents, the drug activity can now be tested on S_{max} and S_{50}. In the first case, the drug activity works to improve (increase or decrease, as appropriate) the maximum limit of the function for a patient as compared to placebo and in the second case, the drug acts to reduce (or increase, again as appropriate) the time to reach the maximum status.

$$S(t) = S_0 + \frac{S_{max}(1 + E(t))t}{S_{50} + t} \tag{21.9}$$

$$S(t) = S_0 + \frac{S_{max}t}{S_{50}(1 + E(t))t} \tag{21.10}$$

A plot of a disease progression model using the E_{max} function is provided in Figure 21.7. In this figure both the treated and untreated curves are provided. Again, the symptomatic improvement is clear because when drug exposure is discontinued, the Disease Severity Score returns to the pretreated trajectory.

A plot of the same E_{max} disease progress model with a disease modifying drug is presented in Figure 21.8. In this case, as was seen previously, the drug effect persists even after the treatment is discontinued.

An example NONMEM control stream for an asymptotic E_{max} model is provided in Table 21.6. In this code, either patients recover following an asymptotic E_{max} func-

FIGURE 21.7 Example profile of asymptotic E_{max} disease progression model with symptomatic drug action.

FIGURE 21.8 Example profile of asymptotic E_{max} disease progression model with disease progression modifying drug action.

tion that maximizes at the predisease score or they decline and do not recover. The use of a mixture model for situations where patients can either improve or worsen is one way to handle a dichotomous situation. However, if a mixture model is used, it is generally preferable to investigate covariates on the probability that patients will worsen. The application of covariate functions to mixture models is presented in Chapter 28 and will not be addressed in this chapter.

Nonzero Asymptotic Function In addition to the exponential and E_{max} models, there are other functions that can describe an asymptotic change in disease score severity over time. Consider a disease severity score such as the Unified Parkinson's

TABLE 21.6 Example NONMEM Code 5: Asymptotic E_{max} Model of Disease Progression

NONMEM Code	Explanation
| ```
$PRED
;Derived values
  TDAY=TIME/24 ; convert hours
    to days
  IF (MIXNUM.EQ.1) THEN
  TV50=THETA(2)
  EC50=TV50*EXP(ETA(4))
  RATE=(EC50*TDAY)
  ELSE
``` | This example is a basic disease progression model with an $E_{max}$ component used to describe only the change in disease severity over time. Consequently, the model can be handled using $PRED rather than the $PK. The control stream is used to describe a situation where patients can either improve or worsen over time, a situation that can follow catastrophic disease or accident. |
| ```
  TVREC=THETA(3)
  TV50=THETA(4)
  REC=TVREC*EXP(ETA(1))
  EC50=TV50*EXP(ETA(2))
  RATE=(REC*TDAY)/(EC50+TDAY)
``` | Here the mixture model is set to describe a steady linear decline when patients worsen over time and an $E_{max}$ model for patients who improve. The selection of the $E_{max}$ model was based on the fact that the disease severity score had a natural maximum. |
| ```
  ENDIF
  EST=MIXEST

  S0=THETA(5)*EXP(ETA(3))
``` | When a mixture model is used for this context, it is necessary to define a common parameter (RATE) that describes the change in severity score over time regardless of the function that is assigned to the patient. |
| ```
;DISEASE SEVERITY SCORE VS
  TIME (days)
  Y=S0+RATE +ERR(1)
``` | As per the definition, this function will either be a linear function or an $E_{max}$ function. RATE is calculated earlier. |
| ```
$MIX ;determines fraction of
  patients responding versus
  declining
  NSPOP=2
  P(1)=THETA(1)
  P(2)=1-P(1)
``` | This is the standard mixture model subroutine for two populations. |
| ```
$THETA (0,.5,1) ; PRDIE
$THETA (0,25,100) ;POPSLOPE
$THETA (-INF, -2,) ; POPIMAX days
$THETA (0,3,100) ;POPIC50
$THETA (0,10,100) ; POPS0 u
$OMEGA BLOCK(3)
  0.1 ;PPVREC
.01 1.5 ;PPVEC50
.01 .01 0.1 ;PPVINT
$OMEGA .1 ;ppvslop
$SIGMA 1 ;SD

$ESTIM MAXEVAL=9990 PRINT=10 SIG=3
METHOD=CONDITIONAL SLOW
$COV
``` | Setting the initial estimates for these models involves evaluating the percentage of subjects in each group and investigating the raw data for initial estimates. In some cases, it is necessary to rerun the model after the first run in order to improve the initial estimates. |

Disease Rating Score (UPDRS), which has a natural maximum value. A maximum total of 199 points is possible on the UPDRS system. A score of 199 represents the worst possible disease status and a score of 0 indicates that the patient has no disability. In this setting, where the scoring system has an inherent maximum value, another exponential model can be considered.

$$S(t) = S_0 e^{-k_{\mathrm{prog}}t} + S_{SS}\left(1 - e^{-k_{\mathrm{prog}}t}\right) \tag{21.11}$$

In this equation, S_{ss} is the maximum limit of the disease severity score, S_0 is the baseline value, and k_{prog} is the rate constant of progression. Symptomatic drug action can be added directly to the function and, as was seen with the $E_{\max}$ model, functions for disease modifying drug action can be implemented on more than one parameter. With this function, the drug may act to lower S_{ss}, or it may act to slow k_{prog}. A plot of this asymptotic function is provided in Figure 21.9. In this figure, the untreated progress is shown along with symptomatic and disease modifying drug effects.

Inverse Bateman   The inverse Bateman function describes transient recovery from a baseline disease severity score, followed by reoccurrence of the disease, and is also useful for describing diseases that exhibit cyclical episodes. The function can also be implemented to describe a transient placebo response when warranted. For example, Holford et al. (25) used the inverse Bateman function to describe the time course of depression in placebo-treated patients. This model was selected for this work in part because of the transient response seen in placebo-treated patients and in part because of the cyclical nature of the disease where patients would be expected to improve and worsen over the course of treatment. Therefore, the first exponential process describes the recovery phase and the second exponential process is used to account for onset of disease in the next episode.

FIGURE 21.9   Example profile of asymptotic disease progression model with all forms of drug action.

FIGURE 21.10 Example profile of inverse bateman disease progression model with symptomatic drug action.

$$HAMD(t) = S_0 - \frac{D_{rec}K_{rec}}{K_{rec} - K_{on}} \exp(K_{on}t - K_{rec}t) \qquad (21.12)$$

In this function, S_0 is the baseline severity score, k_{rec} is the rate constant of recovery, k_{on} is the rate constant of relapse, and D_{rec} is a scale parameter reflecting the amplitude of improvement during recovery. Drug action can be treated again as being symptomatic or disease modifying. In the latter case, the drug can act to decrease the onset rate constant, to increase the offset rate constant, or to increase the amplitude of improvement. A plot of the inverse Bateman disease progress model for an untreated state and a treated state where the drug exhibits symptomatic benefit is provided in Figure 21.10. An example portion of a NONMEM control stream for this model is provided in Table 21.7.

Cosson and Gomeni (26) used a different structural model to describe the time course of depression. These authors used an indirect response model, as this function also provides a means of describing a transient response and ensuing relapse. In the case of depression, the drug would be postulated to transiently alter the rate constants of onset or offset of disease, thereby providing therapeutic benefit. The selection of an indirect response type model seems appropriate since the mechanism of most antidepressant drugs is presumed to be due an inhibitory effect on serotonin reuptake. The model performed well and has a good physiological basis for application. However, for situations where population-based modeling is being conducted, the use of integrated functions can offer shorter run times.

Cyclical Modification of Inverse Bateman Function   A cosine function can be used to describe patterns similar to the exponential and inverse Bateman models and can therefore be used as the function for disease progress. However, this same cosine function can also be used to impose a cyclical modulation on another function.

$$S(t) = Disease\ progress(t) + SADamp \cdot \cos\left(\frac{2\pi}{12} \cdot (Time - Phase)\right) \qquad (21.13)$$

TABLE 21.7   Example NONMEM Code 6: Inverse Bateman Model of Disease Progression with Symptomatic Drug Effect

NONMEM Code	Explanation
```\n$SUBROUTINE ADVAN4\n$PK\nCALLFL=-2\n;PK MODEL\nTVCL=THETA(1)\nPPVCL=ETA(1)\nCL=TVCL*EXP(PPVCL)\nTVV2=THETA(2)\nPPVV2=ETA(2)\nV2=TVV2*EXP(PPVV2)\nTVKA=THETA(3)\nPPVKA=ETA(3)\nKA=TVKA*EXP(PPVKA)\n```	As was done previously, this example control stream calls a two-compartment model in order to use the peripheral compartment as the effect compartment. This provides an empirical means of describing a delay in the onset of effect of drug. The first portion of the control stream then defines the PK parameters.
```\n TVTEQ=THETA(4)\nTEQ=TVTEQ*EXP(ETA(4))\nK=CL/V2\nK23=.001*K;SO THAT LOSS\n TO EFFECT COMT IS TRIVIAL\nK32=LN2/TEQ\nS2=V2\nS3=S2*K23/K32\n```	This section defines the parameters for the effect compartment and provides the parameters in microconstants as is needed for the TRANS1.
```\n$ERROR\nCP=A(2)/S2\nCE=A(3)/S3\n;BEGIN HAMD MODEL\nTVS0=THETA(5)\nS0=TVS0*EXP(ETA(5))\nTVDREC=THETA(6)\nDREC=TVDREC*EXP(ETA(6))\nTVKREC=THETA(7)\nKREC=TVKREC*EXP(ETA(7))\nTVKON=THETA(8)\nKONS=TVKONS*EXP(ETA(8))\n```	Again, because this control stream uses an ADVAN to evaluate the pharmacokinetics, the functions for disease progression and drug action must be defined in the $ERROR routine. Plasma and effect site concentrations are defined as well.
```\n;Derived values\nTDAY=TIME/24; convert hours\n to days\nEXPKO=EXP(-KONS*TNOW)\nEXPKR=EXP(-KREC*TNOW)\nFREC=KREC/(KREC-KONS)\n *(EXPKO-EXPKR)\n```	A derived time scale (days) is done to improve numerical stability and to put the rate constants into more understandable units. The exponential functions for the inverse Bateman function are calculated.
```\n;Drug effect\nTVEM=THETA(9)\nEMAX=TVEM*EXP(ETA(9))\nTVEC=THETA(10)\nEC50=TVEC*EXP(ETA(10))\nEOFF=EMAX*CE/(EC50+CE)\n```	Drug effect parameters are calculated here.
```\nIPRED=F\nIF (DVID .EQ. 1) THEN\nY=CP*EXP(ERR(1)) + ERR(2)\nELSE\n;HAMD SCORE VS TIME (days)\nY=S0-DREC*FREC + EOFF + ERR(3)\nENDIF\n```	Again, the DVID serves as a flag to distinguish concentration records from disease score records. The disease progression model and symptomatic drug effect models are combined under the appropriate DVID flag.

FIGURE 21.11 Example profile of cyclical inverse bateman disease progression model with symptomatic drug action.

In this equation, the disease progression model is evaluated at any time t and the cosine function is added to the overall disease progression model to determine the status. Here, *SADamp* and *Phase* define the amplitude of the underlying cyclical change in disease severity score and the time to the maximum worsening of that score. A plot of an inverse Bateman function with a cyclical component is provided in Figure 21.11.

21.3.2.3 Models for Growth Kinetics

Growth functions comprise an important class of functions that can be applied to various aspects of disease progression. These functions were developed to describe bacterial and tumor growth and are still used for these applications (27–29). However, other markers of disease progression can also, in some cases, be described using a growth function. A simple function that can be used to describe the growth of a response R over time is

$$\frac{dR}{dt} = k_{growth}R - k_{death}R \tag{21.14}$$

In this function, k_{growth} is the growth constant for the response (R) and k_{death} is the rate constant of loss of response. Disease modification can be added to inhibit k_{growth} or to stimulate k_{death}, resulting in enhanced loss of response. Symptomatic drug effect is somewhat less common in these models but can be added as a loss term. A plot of the simple growth function with disease modifying activity is provided in Figure 21.12.

It should be noted that in cases where the cell population is reduced to zero, the model exhibits curative properties as regrowth of the cells cannot occur since the growth rate is a first-order rate depending on an existing population.

FIGURE 21.12 Example profile of simple growth model with disease modifying drug action.

Gompertz Functions   Another series of functions frequently used to describe growth kinetics are the Gompertz functions (30). Gompertz functions describe a rapid initial rate of growth, followed by a slower asymptotic phase of growth to a finite limit. As a consequence of their properties, Gompertz functions have been used to describe the pharmacodynamics of antibacterial agents (31), as well as other systems in which growth kinetics are important. These functions have also been applied to describe the time course of Parkinson's disease (20). A simple Gompertz growth function is shown below, and a plot of this function is provided in Figure 21.13.

$$\frac{dR}{dt} = \beta R(\beta_{max} - R) - [k_{death}R] \tag{21.15}$$

In this equation, R is the response, β is the growth rate constant, β_{max} is the maximum limit for the response, and k_{death} is the rate constant of loss of R. The Gompertz function differs from the simple growth function not only in the changing growth rate, but also in that it has a maximum value above which growth will not occur. In this fashion, the Gompertz function is also an asymptotic function. Like the simple growth function, the model can describe curative behavior if the cell population is brought to zero.

Because this class of functions has been used extensively to model cell growth kinetics, these functions can be modified to describe subpopulations of cells. The equations below describe a Gompertz growth model that has been modified to allow cells to oscillate between a therapeutically sensitive state (R_S) and a resistant state (R_R). The same modification can also be made to the simple growth model (28).

FIGURE 21.13 Example profile of simple Gompertz growth model with disease modifying drug action.

$$\frac{dR_s}{dt} = k_{RS}R_R + \beta R_S(\beta_{\max} - R_S) - \left[k_{SR} + \left(1 + \frac{E_{\max}C_p}{EC50 + C_p} \right) k_{\text{death}} \right] R_S$$

$$\frac{dR_{SR}}{dt} = k_{SR}R_S - k_{RS}R_R \qquad\qquad (21.16)$$

This adaptation of the Gompertz growth function allows for the sensitive cell population to reach zero but will then show delayed regrowth in both sensitive and resistant cell populations as the resistant cell population continues to grow and then transfers to the sensitive cell type.

Transit Models The development of indirect effect models (32) facilitated the evaluation of many disease-mediated processes such as changes in biomarker levels over time. Control streams and example databases for these models are presented elsewhere in Chapters 22 and 23 and will not be covered here. However, as was discussed, when applying these indirect response models, it is possible to increase the delay in response by the addition of extra compartments. The addition of these extra effect compartments also creates additional parameters that need to be estimated, which are often not identifiable.

In 1998, Sun and Jusko (33) investigated the use of transit compartments and gamma distribution functions to describe delayed effects and found the behavior of the transit compartment models acceptable for describing PD observations. In 2000, Friberg et al. (34) utilized the transit model to estimate a mean transit time over the course of white blood cell maturation and to describe the time course of neutropenia following chemotherapy. The model includes a feedback on the synthesis rate for new cell formation based on the observed white cell count. It has since been used to evaluate the neutropenic activity of many other chemotherapeutic agents (35) and can also be used to describe the time course of the effect of administration of exogenous hematopoietic factors (36), as is shown in Figure 21.14.

FIGURE 21.14 Example profile of cell transit model with disease modifying drug action.

FIGURE 21.15 Example profile of modified cell transit model with disease modifying drug action.

A modification of the cell transit model used to describe white blood cells is required when modeling the time course of red blood cell growth such as for anemia (37). In the case of describing the time course of slow growing cells, the cell count can be taken to be the sum of all the component transit compartments, which approximates the lifespan models proposed by Krzyzanski and Jusko (38) and was used to describe the effects of exogenously administered erythropoietin (39). The model proposed by Kryzanski and Jusko was based on the theory that cells have a fixed lifespan. Therefore, the rate of elimination of cells at any given time is dependent on the number of cells formed one lifespan ago. The modified cell transit model described here approximates the lifespan function but does not require that the cell lifespan is fixed. A plot of this modified function is provided in Figure 21.15. This function shows a slow onset and offset of effect of a disease modifying drug action. An example portion of the NONMEM control stream required to implement this model is provided in Table 21.8.

TABLE 21.8 Example NONMEM Code 7: Modified Cell Transit Model Approximating a Cell Lifespan

| ```
$SUBR ADVAN6 TRANS 1 TOL 6
$MODEL
COMP XPO1 ;1 EPO CENTRAL CMT
COMP RBC1 ;2
COMP RBC2 ;3
COMP RBC3 ;4
COMP RBC4 ;5
COMP RBC5 ;6
``` | Because this model cannot be expressed as an integrated function, differential equations must be used. The selection of the appropriate ADVAN (e.g., ADVAN 6, 8, or 9) should be examined carefully. It is not possible to determine a priori if a set of equations is stiff. Therefore, all ADVANs should be tested at least initially. Selection should be based on run times and minimization status. It should be noted that the selection of TOL should be such that it is at least as large as NSIG to improve chances of successful minimization. The compartments are defined here as well. |
| ```
$PK
"FIRST
" COMMON /PRCOMG/ IDUM1, IDUM2,
  IMAX
" INTEGER IDUM1,IDUM2,IMAX
" IMAX=500000
IF (NEWIND.LE.1) THEN
LN2=LOG(2)
D2H=1/24
NRBC=5
ENDIF
``` | $PK is used because this is an ADVAN based model. Verbatim code has been added (as seen by the double quotes starting the lines) to increase the maximum number of iterations. The NEWIND serves to define some values and calculations that are used repeatedly. For such calculations, it is best to do these as little as possible to minimize run time. |
| ```
;RBCTC AND RINNPO
EMXRRBC=THETA(1)*EXP(ETA(1))*D2H
C5XPRBC=THETA(2)*EXP(ETA(2))
HILL=THETA(3)
C5H=C5XPRBC**HILL
; DISEASE PROGRESS
RINNPO=THETA(4)*EXP(ETA(3))
KRBC=1/(THETA(5)*EXP(ETA(4)))*D2H
KRBCN=KRBC*NRBC
;EPO PK
PPVCL=ETA(5)
PPVV1=ETA(6)
PPVTA=ETA(7)
CLXPO=THETA(6)*EXP(PPVCL)*D2H
V1XPO=THETA(7)*EXP(PPVV1)
TABS=THETA(8)/D2H*EXP(PPVTA)
D1=TABS ; ZERO ORDER DURATION
  FOR
EXOGENOUS EPO INPUT
``` | The parameters are defined next. First defined is the rate of hemoglobin formation as a nonlinear function. Then the rate of endogenous erythropoietin is defined.<br><br>Because the cell growth is so slow for some cell types, it is often numerically more stable to define the transfer rate constants as mean residence times (which are larger numbers) and then let the rate constant be 1/MRT as is done here. In addition, certain diseases such as renal failure can result in reduction in loss of synthesis of endogenous epoetin. This loss of synthesis may need to be accounted for in the model.<br><br>For the red blood cell lifespan, several factors must be defined. The first is the effect that both endogenous and exogenous growth factors might have on the cell lifespan and therefore the numbers. |

TABLE 21.8 *Continued*

```
S1=V1XPO
;PD
CNPO=RINNPO/CLXPO
IF (CNPO.GT.0) THEN
CNPH=CNPO**HILL
ELSE
CNPH=0
ENDIF
RRBC0=EMXRRBC*CNPH/(C5H+CNPH)
RBC0=RRBC0/KRBC ; BASELINE RBC
DEPENDENT ON ENDOGENOUS EPO
; INITALIZE RBC CHAIN
F2=RBC0/NRBC
F3=F2
F4=F2
F5=F2
F6=F2
```

The pharmacokinetics of exogenously administered hematopoietic factors are defined and are scaled to days. Again, the use of a longer interval of time (days or weeks) is done for numerical stability because the PD processes are quite slow in some systems. Therefore the PD parameters would be quite small if time were left in hours.

A steady-state concentration of endogenous erythropoietin is assumed and the pharmacokinetics of endogenous and exogenous hematopoietic factors are assumed to be the same. This assumption is not always correct, however, and should be verified.

The baseline synthesis of RBCs is assumed to be based on the endogenous erythropoietin. Again, synthesis is assumed to be nonlinear with a maximum synthesis. It should be noted that this control stream does not have a feedback for increasing the production of endogenous hematopoietic factors.

Finally, the compartments for red blood cells are initialized. The first compartment is set to the ratio of the rate of red blood cells being formed over a fixed normalized value. Because the other compartments are transit compartments, they are assumed to have the same initial conditions at steady state.

```
$DES
DC1XPO=A(1)/V1XPO
;TOTAL EPO CONC IS EXOGENOUS
  PLUS
ENDOGENOUS
DCXPO=DC1XPO+CNPO
IF (DCXPO.GT.0) THEN
DCXPH=DCXPO**HILL
ELSE
DCXPH=0
ENDIF
DRRBC=EMXRRBC*DCXPH/(C5H+DCXPH)
;XPO
DADT(1)=-CLXPO*DC1XPO
;RBC CHAIN
DADT(2)=DRRBC—KRBCN*A(2)
DADT(3)=KRBCN*(A(2) - A(3))
DADT(4)=KRBCN*(A(3) - A(4))
DADT(5)=KRBCN*(A(4) - A(5))
DADT(6)=KRBCN*(A(5) - A(6))
```

Then the differential equations are set up.

The total concentration is assumed to be the sum of endogenous and exogenous erythropoietin. The rate of RBC formation is determined based on a nonlinear function that requires the sum of both.

The first equation is for the pharmacokinetics of exogenous erythropoietin. Then the RBC chain is developed. Note that the rate constant KRBCN is the same for all compartments, as it is for the models of white blood cells.

TABLE 21.8   *Continued*

| ```
$ERROR
C1=A(1)/V1XPO
;TOTAL EPO CONC IS EXOGENOUS
  PLUS
ENDOGENOUS
CXPO=C1 + CNPO
;RBC IS SUM OF ALL AMOUNTS
  IN EACH OF THE
RBC CHAIN COMPARTMENTS
RBC=A(2)+A(3)+A(4)+A(5)+A(6)
``` | The concentration for exogenous erythropoietin is defined again for the $ERROR. It must be defined using a different name than in $DES. <br><br> Again the total concentration of erythropoietin is the sum of endogenous and exogenous concentrations of hematopoietic factors. Here the evaluation for red blood cells differs from the white blood cell transit model. Unlike the white blood cell model, hemoglobin is taken as the sum of all compartments. |
| ```
  IF (DVID.EQ.1) THEN ; EPO OR
ENDOGENOUS EPO CONC
DVXD=2
ADDXPO=THETA(10)
PROXPO=CXPO*THETA(9)
;BSVRUV IS USED TO DESCRIBE
  INDIVIDUAL
RUV
W=SQRT(PROXPO*PROXPO +
ADDXPO*ADDXPO)*EXP(ETA(8))
Y=CXPO + W*ERR(2)
ENDIF
IF (DVID.EQ.2) THEN ; RBC CONC
Y=RBC + ERR(1)
ENDIF
``` | Finally, the functions are evaluated and residual error is estimated. |
| ```
$EST MAX=9990 NSIG=6 PRINT=1
  NOABORT
METHOD=COND INT
``` | Note here that NSIG is 6, which is why TOL had to be set so high. In the initial runs, NSIG and TOL may be lowered to improve the initial estimates. |

21.4   SUMMARY

As was discussed in the introduction to this chapter, one of the main reasons for developing models for disease progression and drug activity is to answer the interesting questions of how the drug works in patients and how to appropriately determine a starting dose. These models provide information that is important for the approval process for the drug and can be used as supportive evidence of activity (40) when only one pivotal trial of efficacy is conducted.

Disease progression models are used to provide information about how to use the drug in practice and can be used in simulation studies to investigate alternative dose regimens. In Figure 21.16, for example, the consequences of alternate dose strategies were investigated for a hypothetical disease modifying agent. The same

FIGURE 21.16 Using the modified cell transit model with disease modifying drug action to evaluate potential dose regimens.

total dose of a hypothetical agent was administered in these nonstochastic simulations to examine increasing dosing intervals. As can be seen, however, the effect of drug is diminished by increasing the dose interval even though the drug exhibits linear pharmacokinetics in this example. Such a finding is consistent with schedule dependence, suggesting that the administered dose in the more frequent interval is providing nearly maximal benefit and that the additional drug given for the longer dose interval is largely inactive due to saturation of the response. Evaluation of the model using simulation is faster and less expensive than running clinical trials.

Other uses for a disease progression model include selection of appropriate designs to assess disease progress and drug action. For example, clinical trial simulation based on a disease progression model was used to design a study to evaluate the effect of levodopa on the progression of Parkinson's disease (41). In particular, the model was used to help ensure that the washout period for active treatment was sufficiently long to determine if levodopa was disease modifying or provided only symptomatic benefit.

The development of disease progression models requires that appropriate assumptions be made and those assumptions should be based on the clinical pharmacology of the drug as well as the disease. In order to ensure that the model assumptions are reasonable, it is important to learn about the disease, the markers used to describe its time course, and other aspects of clinical care of these patients. In short, a team-based approach to the development of disease progression models is generally a good idea. Once developed, these models should be evaluated to determine how well the response surface is estimated and to what extent the model can be used to interpolate and to extrapolate. It is essential that good model building practices and careful evaluation of the model performance be conducted during this exercise.

REFERENCES

1. L. M. Fuller, F. L. Banker, J. J. Butler, J. F. Gamble, and M. P. Sullivan, The natural history of non-Hodgkin's lymphomata stages I and II. *Br J Cancer* **31**(SUPPL 2):270–285 (1975).

2. J. Rosch, R. Antonovic, R. S. Trenouth, S. H. Rahimtoola, D. N. Sim, and C. T. Dotter, The natural history of coronary artery stenosis. A longitudinal angiographic assessment. *Radiology* **119**(3):513–520 (1976).

3. N. H. Holford and L. B. Sheiner, Understanding the dose–effect relationship: clinical application of pharmacokinetic–pharmacodynamic models. *Clin Pharmacokinet* **6**(6): 429–453 (1981).

4. R. C. Griggs, R. T. Moxley, J. R. Mendell, G. M. Fenichel, M. H. Brooke, A. Pestronk, and J. P. Miller, Prednisone in Duchenne dystrophy. A randomized, controlled trial defining the time course and dose response, Clinical Investigation of Duchenne Dystrophy Group. *Arch Neurol* **48**(4):383–388 (1991).

5. D. Barraclough, Rheumatoid arthritis. *Aust Fam Physician* **12**(7):Suppl 2–10 (1983).

6. J. W. Tukey, The future of data analysis. *Ann Math Stat* **33**:13–14 (1962).

7. E. N. Jonsson and L. B. Sheiner, More efficient clinical trials through use of scientific model-based statistical tests. *Clin Pharmacol Ther* **72**(6):603–614 (2002).

8. L. B. Sheiner and D. B. Rubin, Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* **57**(1):6–15 (1995).

9. R. J. Little and D. B. Rubin, Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health* **21**: 121–245 (2000).

10. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

11. P. Bonate, W. R. Gillespie, T. Ludden, D. B. Rubin, L. B. Sheiner, D. Stanski, N. H. G. Holford, M. Hale, H. C. Ko, J.-L. Steimer, and C. C. Peck, *Simulation in Drug Development: Good Practices* `http://cdds.ucsf.edu/research/sddgp.php` 1999.

12. Food and Drug Administration Center for Drug Evaluation and Research, Meeting of the Cardiovascular and Renal Drugs Advisory Committee. Friday, October 20, 2000.

13. N. H. G. Holford, D. R. Mould, and C. Peck, Disease progression models, in *Principles of Clinical Pharmacology*, A. Atkinson (Ed.). Academic Press New York, 2001.

14. P. L. S. Chan and N. H. G. Holford, Drug treatment effects on disease progression. *Annu Rev Pharmacol Toxicol* **41**:625–659 (2001).

15. N. H. Holford and K. E. Peace, Methodologic aspects of a population pharmacodynamic model for cognitive effects in Alzheimer patients treated with tacrine. *Proc Natl Acad Sci USA* **89**:11466–11470 (1992).

16. L. Zhang, S. L. Beal, and L. B. Sheiner, Simultaneous versus sequential analysis for population PK/PD data I: best-case performance. *J Pharmacokinet Pharmacodyn* **30**(6):387–404 (2003).

17. L. B. Sheiner, D. R. Stanski, S. Vozeh, R. D. Miller, and J. Ham, Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to *d*-tubocurarine. *Clin Pharmacol Ther* **25**(3):358–371 (1979).

18. N. H. G. Holford and L. B. Sheiner, Kinetics of pharmacologic response. *Pharmacol Ther* **16**:143–166 (1982).

19. H. C. Kimko, S. S. Reele, N. H. Holford, and C. C. Peck, Prediction of the outcome of a phase 3 clinical trial of an antischizophrenic agent (quetiapine fumarate) by simulation

with a population pharmacokinetic and pharmacodynamic model. *Clin Pharmacol Ther* **68**(5):568–577 (2000).

20. N. Holford, P. Chan, and J. Nutt, Disease progression in Parkinson's disease—evidence for protective effects of drug treatment. Abstract 478, Population Analysis Group in Europe, Barcelona, Spain, 2004.

21. S. Pors Nielsen, O. Barenholdt, F. Hermansen, and N. Munk-Jensen, Magnitude and pattern of skeletal response to long term continuous and cyclic sequential oestrogen/ progestin treatment. *Br J Obstet Gynaecol* **101**(4):319–324 (1994).

22. G. Pillai, R. Gieschke, T. Goggin, P. Jacqmin, R. C. Schimmer, and J. L. Steimer, A semimechanistic and mechanistic population PK–PD model for biomarker response to ibandronate, a new bisphosphonate for the treatment of osteoporosis. *Br J Clin Pharmacol* **58**(6):618–631 (2004).

23. B. J. Anderson, N. H. G. Holford, G. A. Woollard, S. Kanagasundaram, and M. Mahade- van, Perioperative pharmacodynamics of acetaminophen analgesia. *Children Anesthesi- ology* **90**(2):411–421 (1999).

24. T. Taylor, D. R. Mould, and L. J. Benincosa, Population pharmacokinetics and pharma- codynamics of CP-101,606 following a 72-hour infusion in acute ischemic stroke patients. American Society for Clinical Pharmacology and Therapeutics, Annual Meeting, Miami, FL, March 2004.

25. N. Holford, J. Li, L. Benincosa, and M. Birath, Population disease progress models for the time course of HAMD score in depressed patients receiving placebo in anti- depressant clinical trials. Abstract 311, Population Analysis Group in Europe, Paris, France, 2002.

26. V. Cosson and R. Gomeni, Modelling placebo response in depression using a mechanis- tic longitudinal model approach. Abstract 818, Population Analysis Group in Europe, Pamplona, Spain, 2005.

27. W. Jusko, Pharmacodynamics of chemotherapeutic effects: dose–time–response relation- ships for phase-nonspecific agents. *J Pharm Sci* **60**:892–895 (1971).

28. W. Jusko, Pharmacodynamic model for cell-cycle-specific chemotherapeutic agents. *J Pharmacokinet Biopharm* **1**:175–200 (1973).

29. J. Zhi, C. H. Nightingale, and R. Quintilliani, A pharmacodynamic model for the activ- ity of antibiotics against microorganisms under nonsaturable conditions. *J Pharm Sci* **75**:1063–1067 (1986).

30. P. Royston, A useful monotonic non-linear model with applications in medicine and epidemiology. *Stat Med* **19**:2053–2066 (2000).

31. Y. Yano, T. Oguma, H. Nagata, and S. Sasaki, Application of a logistic growth model to pharmacodynamic analysis of in vitro bacteriocidal kinetics. *J Pharm Sci* **87**:1177–1183 (1998).

32. N. L. Dayneka, V. Garg, and W. Jusko, Comparison of four basic models of indirect pharmacodynamic response. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

33. Y.-N. Sun and W. Jusko, Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* **87**:732–737 (1998).

34. L. E. Friberg, A. Freijs, M. Sandstrom, and M. O. Karlsson, Semiphysiological model for the time course of leukocytes after varying schedules of 5-fluorouracil in rats. *J Pharma- col Exp Ther* **295**:734–740 (2002).

35. E. D. Lobo and J. P. Balthasar, Pharmacodynamic modeling of chemotherapeutic effects: application of a transit compartment model to characterize methotrexate effects in vitro. *Am Assoc Pharm Sci Pharm Sci* **4**(4):42 (2002).

36. D. R. Mould, PKPD models for white cell responses to hematopoietic stimulation with and without chemotherapy. American Society of Clinical Pharmacology and Therapeutics Annual Meeting, Orlando, FL, March 2–5, 2005.

37. N. H. G. Holford, Pharmacokinetic Pharmacodynamic models for red cell responses to hematopoietic stimulation with and without chemotherapy. American Society of Clinical Pharmacology and Therapeutics Annual Meeting, Orlando, FL, March 2–5, 2005.

38. W. Krzyzanski and W. J. Jusko, Multiple-pool cell lifespan model of hematologic effects of anticancer agents. *J Pharmacokinet Pharmacodyn* **29**(4):311–337 (2002).

39. W. Krzyzanski, W. J. Jusko, M. C. Wacholtz, N. Minton, and W. K. Cheung, Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after multiple subcutaneous doses in healthy subjects. *Eur J Pharm Sci* **26**(3–4):295–306 (2005).

40. C. C. Peck, D. B. Rubin, and L. B. Sheiner, Hypothesis: a single clinical trial plus causal evidence of effectiveness is sufficient for drug approval. *Clin Pharmacol Ther* **73**(6):481–490 (2003).

41. P. L. S. Chan, N. H. G. Holford, and J. Nutt, Application of clinical trial simulation to evaluate the ELLDOPA trial design. Movement Disorders Society Conference, Miami, FL, November 2002.

APPENDIX 21.1 EXAMPLE FOR THE SIMPLE LINEAR DISEASE PROGRESSION MODEL

```
$PROB PLACEBO AND ACTIVE PK AND OUTCOME
$INPUT READ NOTE=DROP ID DATE=DROP TIME DV DVID
AMT ADDL II MDV EVID
$DATA ..\CH21EXAMPLE01.csv WIDE IGNORE #
;The data file is the Excel spreadsheet named CH21EXAMPLE01.csv
$SUBROUTINE ADVAN4
$PK
  CALLFL=-2
  IF (NEWIND.LE.1) THEN
    D2H=24
    LN2=LOG(2)
    TWOPI=2*3.141592654
    DOSE=0
  ENDIF
  IF (AMT .NE. 0) DOSE = AMT

;PK MODEL

  TVCL=THETA(1)
  CL=TVCL*EXP(ETA(1))

  TVV2=THETA(2)
  V2=TVV2*EXP(ETA(2))

  TVTEQ=THETA(4)
  TEQ=TVTEQ*EXP(ETA(8))
```

```
  K=CL/V2
  K23=.001*K ;SO THAT LOSS TO EFFECT COMT IS TRIVIAL
  K32=LN2/TEQ
  S2=V2
  S3=S2*K23/K32

  TVKA=THETA(3)
    KA=TVKA*EXP(ETA(3))

$ERROR
  CP=A(2)/S2
  CE=A(3)/S3

;BEGIN HAMD MODEL

  TVS0=THETA(7)
  TVSLOP=THETA(8)

  S0=TVS0*EXP(ETA(4))
  SLOPE=TVSLOP*EXP(ETA(5))

;DERIVED VALUES
  TDAY=TIME/D2H ; CONVERT HOURS TO DAYS
  REC=S0-SLOPE*TDAY

;DRUG EFFECT
  TVEM=THETA(5)
  EMAX=TVEM*EXP(ETA(6))
  TVEC=THETA(6)
  EC50=TVEC*EXP(ETA(7))
  EOFF=EMAX*CE/(EC50+CE)

IPRED=F
IF (DVID .EQ. 1) THEN
    Y=CP*EXP(ERR(1)) + ERR(2)
ELSE
;HAMD SCORE VS TIME (DAYS)
    Y=REC+EOFF + ERR(3)
ENDIF

;PARAMS FOR PK
$THETA (20) FIX; CL
$THETA (100) FIX ; V2
$THETA (.1) FIX ; KA
$THETA (96) ; TEQ

;DRUG EFFECT PARAMS
$THETA (6.79) ;EMAX
$THETA (80,) ;EC50

;PARAMS FOR HAMD MODEL
$THETA (24) ; POPS0 DAYS
$THETA (0.31) ; POPSLO
```

```
;VARIANCE FOR PK
$OMEGA .1 FIX; PPVCL
$OMEGA .1 FIX; PPVV2
$OMEGA .1 FIX ; PPVKA

;VARIANCE FOR HAMD
$OMEGA 0.2 ;PPVS0
$OMEGA 0.5 ;PPVSLO

;VARIANCE FOR DRUG EFFECT
$OMEGA 0.6 ;PPVEM
$OMEGA 0.8 ;PPVEC
$OMEGA .5 ;PPVTEQ

;RESIDUAL FOR PK
$SIGMA 0.1 FIX ; CVCP
$SIGMA 1 FIX ; SDCP

;RESIDUAL FOR HAMD
$SIGMA 1 ;SD

$EST MAX=9990 SIG=3 NOABORT PRINT=1
METHOD=COND
MSFO=example_run.msf

$TABLE ID TIME CP CE EMAX EC50 TEQ
EOFF DVID Y
NOPRINT ONEHEADER FILE=example_run.fit
```

Mechanistic Pharmacokinetic/ Pharmacodynamic Models I

VARUN GARG and ARIYA KHUNVICHAI

22.1   INTRODUCTION

Over the last several years, pharmacokinetic/pharmacodynamic (PK/PD) modeling has gained increasing importance in its application to drug development. A major reason for this is the progress made in the identification of pharmacodynamic markers (i.e., "biomarkers") of drug response that can be measured easily and the development of mechanism-based PK/PD models that allow one to quantify and predict drug effects under different dosages or physiologic conditions. In some cases, mechanism-based modeling has even provided useful insights into the possible mode of action of drugs (1, 2).

One of the earliest examples of a mechanism-based PK/PD model was the model developed for the indirect anticoagulant effect of warfarin (3). The model accounted for the delay observed between the plasma concentrations of warfarin and the anticoagulant effect and described the linear relationship between warfarin concentrations and its direct effect on the synthesis rate of vitamin K-dependent clotting factors.

Jusko (4, 5) described pharmacodynamic (PD) models for cell proliferation and irreversible effects of chemotherapeutic agents. These models and others using similar concepts are reviewed in Chapter 23.

Dayneka et al. (6) characterized the four basic indirect response models—the first set of mechanistic PK/PD models to describe a diverse array of drug responses. Shortly thereafter, the importance (7) and initial applications (8, 9) of these models were described using data obtained from the literature. Since then, numerous applications and advancements of these models have been published.

The development of mechanistic models for viral dynamics in the last decade has played an important role in the understanding of pathogenesis of viral diseases such as human immunodeficiency virus (HIV) (10–12), hepatitis B virus (HBV)

(13, 14), and hepatitis C virus (HCV) (1, 15–17). However, PK concepts have been incorporated into these models only in last few years (2, 17, 18).

In this chapter, we review the basic concepts and their applications of these models. The implementation of these models using WinNonlin® (Pharsight Corporation, Mountain View, CA) and/or NONMEM® (GloboMax LLC, Ellicott City, MD) is provided in the appendix to this chapter.

22.2 INDIRECT RESPONSE MODELS

The term "indirect response" refers to a PD response that is produced by a drug's action on the production or dissipation of endogenous factors that affect the response. Thus, the measured response is indirectly related to the direct effect produced by the drug at the site of action: for example, the reduction in pain by the inhibitory action of nonsteroidal anti-inflammatory drugs on the production of endogenous pain mediators.

22.2.1 Basic Models of Indirect Response

In the simplest scheme, the rate of change of the response when no drug is present is described by the following equation (6):

$$\frac{dR}{dt} = k_{in} - k_{out}R \tag{22.1}$$

where k_{in} represents the zero-order rate constant for production of response, R, and k_{out} is the first-order rate constant for the loss of response variable. The response variable R may be a directly measured entity or an observed response, which is immediately proportional to the concentration of R. As the system is assumed to be stationary for these models, the response variable (R) begins at a predetermined baseline value (R_0), changes with time following drug administration, and eventually returns back to R_0. Thus, $k_{in} = k_{out}R_0$.

In the basic models, four possible permutations of the response have been characterized (6) involving either the inhibition or stimulation of k_{in} or k_{out} to account for the most commonly expected types of responses (see Figure 22.1). The inhibitory function, $I(t)$, and the stimulatory function, $S(t)$, can be described as

FIGURE 22.1 Schematic of the four basic models of indirect response. The solid bars represent inhibition and the open bars represent stimulation of the input and output functions. (Adapted from Ref. 50.)

$$I(t) = 1 - \frac{C}{C + IC_{50}} \tag{22.2}$$

$$S(t) = 1 + \frac{E_{max}C}{C + EC_{50}} \tag{22.3}$$

where C represents the plasma concentration of the drug as a function of time, IC_{50} (or EC_{50}) is the drug concentration that produces 50% of maximum inhibition (or stimulation, in case of EC_{50}) achieved at the site of effect, and E_{max} represents the maximum effect attributed to the drug. A Hill coefficient, γ, can be added for modeling the sigmoidicity, where needed.

In Eq. (22.2), it is assumed that at high doses of the drug, a complete inhibition of k_{in} or k_{out} is achieved. For some drugs, k_{in} or k_{out} may not be completely inhibited at high doses. In this case, a modification of models I and II, that includes the maximum inhibition (I_{max}), can be made (19, 20):

$$I(t) = 1 - \frac{I_{max}C}{C + IC_{50}} \tag{22.4}$$

where $0 < I_{max} \leq 1$. The four models are described next.

22.2.1.1 Model I: Inhibition of the Production of Response Mediator

$$\frac{dR}{dt} = k_{in}I(t) - k_{out}R \tag{22.5}$$

Model I has been applied to a wide variety of drug responses, such as the reduction of fever (8, 20) or pain (19) by anti-inflammatory drugs, anticoagulant action of warfarin (3, 9), reduction in blood sorbitol levels by inhibitors of aldose reductase (21), cortisol suppressive effects of corticosteroids (22), luteinizing hormone suppression by the synthetic hormone cetrorelix (23), reduction in the levels of tumoral phospho-EGFR (epidermal growth factor receptor) by cetuximab (24), inhibition of dihydrotestosterone (25), the suppression of T-lymphocyte influx into the blood by corticosteroids (26), and the acid-inhibitory effects of H_2-receptor antagonists (27).

22.2.1.2 Model II: Inhibition of the Dissipation of Response Mediator

$$\frac{dR}{dt} = k_{in} - k_{out}I(t)R \tag{22.6}$$

This model has been applied to the inhibition of cholinesterase (9), the inhibition of water reabsorption by loop diuretics such as furosemide (9), tryptophan-mediated increase in hepatic activity of tryptophan pyrrolase (28), and the accumulation of lymphocytes in the peripheral lymphoid tissues by prednisolone (26).

22.2.1.3 Model III: Stimulation of the Production of Response Mediator

$$\frac{dR}{dt} = k_{in}S(t) - k_{out}R \tag{22.7}$$

This model has been applied to the induction of MX protein by interferon alfa-2a (29), bronchodilatory effect of β_2-adrenergic agonists (9), induction of prolactin secretion by H_2-receptor antagonists (9) or by dopamine antagonists (30, 31), hydrocortisone-mediated stimulation of hepatic tryptophan pyrrolase activity (28), stimulation of the secretion of growth hormone by growth hormone releasing peptides (32), induction of neutrophil production by drugs such as prednisolone (33), and the stimulation of the production rate of CD34[+]cells by CXCR4 antagonist (34).

22.2.1.4   Model IV: Stimulation of the Dissipation of Response Mediator

$$\frac{dR}{dt} = k_{in} - k_{out}S(t)R \tag{22.8}$$

Model applications include terbutaline's effect on lowering plasma potassium levels (9, 35) and stimulation of the factors controlling heat loss by the antipyretic effect of nonsteroidal anti-inflammatory drugs (36).

Figure 22.2 shows typical response–time profiles obtained from simulation for increasing doses of a hypothetical drug. In each of the models, a slow increase or decrease of response is seen, until the maximum response (R_{max}) is reached. This is followed by a gradual return of the response to baseline when the drug is discontinued. The maximum response lags behind the maximum plasma concentration of the drug (C_{max}) and the drug response lasts beyond the presence of effective concentrations in the plasma. The maximum response and the time of maximal response are dependent on the dose, I_{max} (or E_{max}) and IC_{50} (or EC_{50}). The basic properties and signature patterns of each of the four models were examined by Sharma and Jusko (37, 38) and Krzyzanski and Jusko (39) and may be helpful in experimental designs and in assigning appropriate models to the data.

When the hypothetical effect-compartment model (40) is applied to such data, the model is usually quite good when fitted to the drug response at individual doses; however, the model parameters are often dose-dependent and the model does a poor job when fitted to data from several doses simultaneously (6, 28). The utility of the hypothetical effect-compartment model in predicting responses at other doses, therefore, is limited.

22.2.2   Natural Cell Lifespan Models

Indirect response models have been extended to drugs that alter the generation of natural cells (41). Unlike cancerous cells and embryonic stem cells, the lifespan of primary human cells is finite. Thus, cells live for a specific duration known as the cell lifespan and then undergo apoptosis (programmed cell death). Cell lifespan models assume that, for a given cell type, each cell lives for the same period of time T_R and then disappears (Figure 22.3). Thus, cells are produced at a zero-order rate k_{in} and are lost at the same rate, k_{in}, but T_R units of time later:

$$\frac{dR}{dt} = k_{in}(t) - k_{in}(t - T_R) \tag{22.9}$$

FIGURE 22.2 Simulations of the response variables with time after a single oral dose of 1 mg (solid line), 10 mg (dashed line), 100 mg (shaded line), or 1000 mg (solid-dashed line) using the four basic indirect response models. The PK parameter values used were $k_a = 0.8$ h^{-1}, $k_{el} = 0.4 h^{-1}$, $F = 1$, $V = 30 L$; the PD parameters were $IC_{50} = 10$ ng/mL, $E_{max} = 1$, $k_{in} = 10$ units/h, $k_{out} = 0.1 h^{-1}$, $R_0 = 100$ units ($k_{in} = k_{out}R_0$). (Adapted from Ref. 6.)

FIGURE 22.3 Schematic of the lifespan model.

At baseline,

$$R_0 = k_{in}T_R \tag{22.10}$$

For a drug that stimulates cell production, we apply the stimulatory function $S(t)$:

$$\frac{dR}{dt} = k_{in}S(t) - k_{in}S(t - T_R) \tag{22.11}$$

where $S(t)$ and $S(t - T_R)$ are given by the following equations:

$$S(t) = 1 + \frac{S_{max}C(t)^\gamma}{C(t)^\gamma + SC_{50}^\gamma} \tag{22.12}$$

$$S(t - T_R) = 1 + \frac{S_{max}C(t - T_R)^\gamma}{C(t - T_R)^\gamma + SC_{50}^\gamma} \tag{22.13}$$

These models have been applied to describe the stimulatory effects of hematopoietic growth factors such as granulocyte colony stimulating factor on neutrophils, thrombopoietin on platelets, and erythropoietin on reticulocytes in blood (41). For the myelosuppressive effects of anticancer drugs, a two-compartment indirect model (42) or a multiple-pool cell lifespan model has been described (43).

Several other advancements/modifications in the basic models for indirect response have been made. Zuideveld et al. (44, 45) described a model for hypothermic effect that incorporated a set-point temperature that is decreased by 5-HT$_{1A}$ receptor agonists. Lima et al. (35) described the stimulation of glucose production and, consequently, the stimulation of insulin production by β_2-adrenergic agonists, using a positive-feedback system. Fasanmade and Jusko (46) proposed a model for formation of methemoglobin by antimalarials that used reactive metabolites as the biophase. Zamboni et al. (47) described the time course of topotecan-induced neutropenia with a model for inhibition of stem cell production linked to an effect compartment. Krzyzanski and Jusko (48) and Li et al. (49) describe a peripheral response pool for responses with multicompartmental distribution. Further advancements, including the integration of irreversible effects, transduction processes, and tolerance and rebound phenomenon are described in Chapter 23.

22.2.3 Limitations of Indirect Response Models

Although indirect response models can be applied to characterize the pharmacodynamics of numerous drugs, some practical limitations exist in their applications:

1. An understanding of the pathophysiology of the disease is required for building the proper model.
2. The models operate best when the response variable immediately reflects the production and loss processes.
3. Indirect response models require differential equations and numerical integration algorithms to describe the nonlinear inhibition or stimulation. Partially integrated solutions for these models have been developed (50, 51), which allow qualitative examination of the relationships between response

and dose or the various PD model parameters. The use of model I with an empirical solution (i.e., without numerical integration) has been suggested (52); however, Krzyzanski and Jusko (53) demonstrated that this approach may result in errors in the estimated parameters and is of limited value.

4. An understanding of the progression of the response in the absence of drug is essential for the proper application of the model. For example, endogenous cortisol (24, 54–56) and osteocalcin (57) levels in plasma follow a circadian pattern that can be fitted with a cosine function.

22.3   VIRAL DYNAMIC MODELS

As mentioned earlier, several mechanistic models for viral dynamics during antiviral treatment for blood-borne viruses including hepatitis C virus (HCV), hepatitis B virus (HBV), and human immunodeficiency virus (HIV) have been developed. The general framework of these models is analogous to the indirect response models with some important differences that will become apparent in the following section.

A general schematic of these models is shown in Figure 22.4. Briefly, there are three main compartments in the model: the target cells (T), the infected cells (I), and the virus (V). Target cells are synthesized by a zero-order rate (s), are infected with a de novo infection rate, βVT, or die with a death rate constant T_d (see Eq. (22.14)). Here β is a second-order rate constant. Target cells are productively infected by the virus and these infected cells are eliminated with a first-order rate constant, δ (see Eq. (22.15)). Productively infected cells release new virus with a first-order rate constant, p, and free virus particles are cleared with a rate constant c (see Eq. (22.16)).

$$\frac{dT}{dt} = s - T_d T - \beta VT \tag{22.14}$$

FIGURE 22.4   Schematic of the viral dynamic model.

$$\frac{dI}{dt} = \beta VT - \delta I \tag{22.15}$$

$$\frac{dV}{dt} = pI - cV \tag{22.16}$$

The number of virions (viral load) in the blood is currently the major endpoint that is used to evaluate the efficacy of antiviral drugs. Therefore, mathematical modeling of the viral dynamics early in the treatment can provide a good indication of the effectiveness of the drug and, in some cases (e.g., in the treatment of hepatitis C), the duration of treatment required. In the future, we may see the use of viral dynamic models in optimizing drug combinations.

22.3.1   Modeling Hepatitis C Viral Dynamics During Treatment

Although a number of other agents are undergoing clinical trials for the treatment for hepatitis C virus (HCV), only ribavirin and interferon alpha are approved. While the exact mechanism of action of these agents is unclear, viral dynamic modeling suggests that interferon alpha acts by decreasing the production rate of new HCV virions from infected cells rather than blocking de novo infection (1). Ribavirin, by itself, has negligible effect on HCV viral load (58). However, when combined with interferon alpha, it has a synergistic effect and improves treatment outcome (2). Modeling of the additional effect of ribavirin in combination with interferon alpha (2) suggests that ribavirin decreases HCV infectivity and increases the proportion of noninfectious virus. During treatment with a combination of ribavirin and interferon, Eqs. (22.14)–(22.16) can be modified to Eqs. (22.17–22.20):

$$\frac{dT}{dt} = s - T_d T - \beta V_I T \tag{22.17}$$

$$\frac{dI}{dt} = \beta V_I T - \delta I \tag{22.18}$$

$$\frac{dV_I}{dt} = (1 - \rho)(1 - \varepsilon)pI - cV_I \tag{22.19}$$

$$\frac{dV_{NI}}{dt} = \rho(1 - \varepsilon)pI - cV_{NI} \tag{22.20}$$

where ε and ρ represent the effectiveness of interferon alpha and ribavirin, respectively ($0 \leq \varepsilon, \rho \leq 1$). For interferon alpha monotherapy, $\rho = 0$. The total number of virions (V) is the sum of the infectious virions (V_I) and noninfectious virions (V_{NI}).

During interferon monotherapy ($\rho = 0$), the viral load declines rapidly during the first 48 hours and more slowly thereafter (1). In theory, if the effectiveness of treatment is 100% ($\varepsilon = 1$), the viral load will decline monoexponentially with a slope equal to its clearance rate (cV). However, in practice, ε is always less than 1 and the rate of viral load decline is biexponential. The analysis of viral decline (Eqs. (22.17)–(22.20)) indicates that the first slope is determined by the effectiveness of interferon (ε) and the free virus clearance rate (c) and the second slope is determined by the productively infected cell clearance rate constant (δ) and the effectiveness of drugs (ε) as shown in Figure 22.5.

FIGURE 22.5 Simulation of HCV viral dynamic profiles during interferon alpha mono-therapy using Eqs. (22.17)–(22.20). For (A), parameters were $\rho = 0$, $c = 5\,\mathrm{d}^{-1}$, $\delta = 0.24\,\mathrm{d}^{-1}$, $\beta = 3 \times 10^{-7}$ (virion per mL)$^{-1}$ per day, $p = 100$ virions/mL/cell/day, and varying $\varepsilon = 1$, 0.8, and 0.5 for dashed, dotted, and solid lines, respectively. For (B), parameters were $\rho = 0$, $c = 5\,\mathrm{d}^{-1}$, $\varepsilon = 0.9$, $\beta = 3 \times 10^{-7}$ (virion per mL)$^{-1}$ per day, $p = 100$ virions/mL/cell/day, and varying $\delta = 0.24$ and $0.12\,\mathrm{d}^{-1}$ for dashed and solid lines, respectively. For (C), parameters were $\rho = 0$, $\varepsilon = 0.9$, $\delta = 0.24\,\mathrm{d}^{-1}$, $\beta = 3 \times 10^{-7}$ (virion per mL)$^{-1}$ per day, $p = 100$ virions/mL/cell/day, and varying $c = 5$ and 15 for solid and dashed lines, respectively.

The results of simulations of a combination treatment with interferon alpha and ribavirin are shown in Figure 22.6. Using such simulations, Dixit et al. (2) showed that ribavirin improves the second phase of viral decline when the effectiveness of interferon is low.

FIGURE 22.6 Simulation of HCV viral dynamic profiles during combination treatment with interferon alpha and ribavirin using Eqs. (22.17)–(22.20), where $\varepsilon = 0.9$ and 0.5 for the lower two lines (**C** and **D**) and upper two lines (**A** and **B**), respectively, with $\rho = 1$ for **B** and **D**, respectively, and 0.1 for **A** and **C**, respectively. Other parameters are $c = 5\,d^{-1}$, $\delta = 0.24\,d^{-1}$, $\beta = 3 \times 10^{-7}$ (virion per mL)$^{-1}$ per day, and $p = 100$ virions/mL/cell/day.

22.3.2   Modeling for HBV During Treatments

The treatment of HBV with lamivudine (LMV) and famciclovir (FCV) therapy also exhibits a biphasic decline in viral load with an initial rapid decline during the first 2 days representing both the clearance of free virus rate (cV) and the effectiveness of drugs (ε) followed by a slower decline representing the clearance of infected cells (δ) (13, 59, 60). Lewin et al. (60) and Nowak et al. (13) proposed that during LMV and FCV therapy, the production rate of new virus (pI) was decreased by a factor $(1 - \varepsilon)$ and de novo infection rate (βVT) was decreased by a factor $(1 - \eta)$. Therefore, Eqs. (22.14)–(22.16) can be modified as follows:

$$\frac{dT}{dt} = s - T_d T - (1 - \eta)\beta VT \tag{22.21}$$

$$\frac{dI}{dt} = (1 - \eta)\beta VT - \delta I \tag{22.22}$$

$$\frac{dV}{dt} = (1 - \varepsilon)pI - cV \tag{22.23}$$

These models suggest that HBV viral dynamic profiles are similar to HCV viral dynamic profiles.

22.3.3   Modeling for HIV During Treatments

Viral dynamic modeling has been described for the effects of two classes of anti-HIV drugs, namely, reverse-transcriptase inhibitors (RTIs) that prevent infection of new cells and protease inhibitors (PIs) that decrease production of infectious virions by blocking the release of virions from infected cells, leading to the production of

noninfectious virions. In the presence of drug in the system, Eqs. (22.14)–(22.17) are modified to Eqs. (22.24)–(22.27):

$$\frac{dT}{dt} = s - T_d T - (1 - \varepsilon_{RT})\beta V_I T \tag{22.24}$$

$$\frac{dI}{dt} = (1 - \varepsilon_{RT})\beta V_I T - \delta I \tag{22.25}$$

$$\frac{dV_I}{dt} = (1 - \varepsilon_{PI})PI - cV_I \tag{22.26}$$

$$\frac{dV_{NI}}{dt} = \varepsilon_{PI} pI - cV_{NI} \tag{22.27}$$

ε_{RT} and ε_{PI} are the effectiveness of RTI and PI, and V_I and V_{NI} are infectious and noninfectious virions, respectively. HIV viral dynamic curves are the same pattern as HCV and HBV viral kinetic curves (61, 62). Perelson (62) theorized that HIV-1 was cleared from chronically infected subjects at a rapid rate, with a half-life of 6 hours or less, whereas the free virus clearance rate constant (c) varied between 9.1 per day and 36 per day (59). Using these parameters, simulations were performed using Eqs. (22.24)–(22.27), which indicate that the first slope of rapid decline was dependent on the elimination rate of the productively infected cells and the effectiveness of drugs (see Figure 22.7). To explain the results from long-term combination treatment, Perelson et al. (61) and Ding and Wu (63) assumed that there are two major HIV-infected cell compartments— productively infected cells (I_I) and long-lived infected cells (I_L). Other compartments, such as latently infected cells, may also exist, but these compartments cannot be identified from plasma viral load measurements. Without treatment, target cells (T) get infected from infected virions with the rate $\beta V_I T$. The proportion of productively infected cells and long-lived infected cells are α_1 and α_2, respectively. The average rate of virus production per cell, p, is given by $N\delta$, where δ is the productively infected cell clearance rate constant, N is the number of new virions produced per infected cell, and η_0 represents the proportion of noninfectious virus in the total virus before the treatments. Other parameters were kept the same as before. The basic viral dynamic model before treatment can then be written as follows (61, 63, 64):

$$\frac{dI_I}{dt} = \alpha_1 \beta V_I T - \delta_1 I_I \tag{22.28}$$

$$\frac{dI_L}{dt} = \alpha_2 \beta V_I T - \delta_L I_L \tag{22.29}$$

$$\frac{dV_I}{dt} = (1 - \eta_0)(N_I \delta_I I_I + N_L \delta_L I_L) - cV_I \tag{22.30}$$

$$\frac{dV_{NI}}{dt} = \eta_0(N_I \delta_I I_I + N_L \delta_L I_L) - cV_{NI} \tag{22.31}$$

Reverse transcriptase inhibitors reduce the production of de novo infection by factors $(1 - \gamma_1)$ and $(1 - \gamma_2)$ in the two infected cell compartments, I_I and I_L,

FIGURE 22.7 Simulation of HIV-1 viral dynamic profiles during combination of protease inhibitors and reverse transcriptase inhibitors using Eqs. (22.24)–(22.27) (not incorporating the long-lived infected cells) where $\delta_I = 4\,d^{-1}$ with twofold difference in ε (A), δ_I (B), and c (C).

respectively. Protease inhibitors reduce the production of the infectious virus from productively infected cells and long-lived infected cells by factors $(1 - \eta_1)$ and $(1 - \eta_2)$, respectively. Therefore, the combination of these antiviral drugs can be described by the following equations:

$$\frac{dI_I}{dt} = (1 - \gamma_1)\alpha_1\beta V_I T - \delta_1 I_I \qquad (22.32)$$

$$\frac{dI_L}{dt} = (1 - \gamma_2)\alpha_2\beta V_I T - \delta_T I_L \qquad (22.33)$$

FIGURE 22.8 Simulations of HIV-1 viral dynamic profiles during combination of protease inhibitors and reverse transcriptase inhibitors using Eqs. (22.32)–(22.35) (incorporates long-lived infected cells), where $\delta_I = 4\,\mathrm{d}^{-1}$, $\delta_L = 0.4\,\mathrm{d}^{-1}$ for a solid line and $\delta_I = 4\,\mathrm{d}^{-1}$, $\delta_L = 0.8\,\mathrm{d}^{-1}$ for a dashed line.

$$\frac{dV_I}{dt} = (1 - \eta_0)((1 - \eta_1)N_I\delta_I I_I + (1 - \eta_2)N_L\delta_L I_L) - cV_I \tag{22.34}$$

$$\frac{dV_{NI}}{dt} = [\eta_0 + (1 - \eta_0)\eta_2]N_L\delta_L I_L + [\eta_0 + (1 - \eta_0)\eta_I]N_I\delta_I I_I - cV_{NI} \tag{22.35}$$

Differences in the lifespans of infected cells resulting in several phases of HIV kinetics may exist after treatment. Simulations using these equations (shown in Figure 22.8) indicate that the first slope is determined mainly by elimination of productively infected cells and the second slope is determined by the loss of long-lived infected cells.

22.3.4 Limitation of Viral Dynamic Models

The current models do not describe the development of resistance and/or the lack of response. Limitations exist in identifying infectious versus noninfectious or uninfected target cells. As mentioned earlier, these models do not account for the changes in the pharmacokinetics of the drugs and assume that their effectiveness is constant during treatment. Recently, however, it has been proposed that the effectiveness, ε, be linked to the drug concentrations and IC_{50} as shown by Eq. (22.36) (17):

$$\varepsilon(t) = \frac{IC_{50}^n}{IC_{50}^n + C^n} \tag{22.36}$$

22.4   SUMMARY

Mechanism-based models have gained importance in drug development. Two general classes of mechanistic models—indirect response models and viral dynamic models—are described here. Indirect response models characterize a wide variety of pharmacologic response and have been described here along with their applications, extension/advancements, and limitations. Viral dynamic models are mechanistic models that are evolving to incorporate the pharmacokinetics of the various classes of drugs becoming available for treatment for important diseases such as HIV and hepatitis B or C. Together, these models along with those presented in Chapter 23 provide the current "state of the art" in mechanism-based PK/PD modeling.

REFERENCES

1. A. U. Neuman, N. P. Lam, H. Dahari, D. R. Gretch, T. E. Wiley, T. J. Layden, and A. S. Perelson, Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-α therapy. *Science* **282**:103–107 (1998).

2. N. M. Dixit, J. E. Layden-Almer, T. J. Layden, and A. S. Perelson, Modelling how ribavirin improves interferon response rates in hepatitis c virus infection. *Nature* **432**:922–924 (2004).

3. R. Nagashima, R. A. ÓReilly, and G. Levy, Kinetics of pharmacologic effects in man: the anticoagulant action of warfarin. *Clin Pharmaco Ther* **10**:22–35 (1969).

4. W. J. Jusko, Pharmacodynamics of chemotherapeutic effects: dose–time–response relationships for phase-nonspecific agents. *J Pharm Sci* **60**:892–895 (1971).

5. W. J. Jusko, A pharmacodynamic model for cell-cycle-specific chemotherapeutic agents. *J Pharmacokinet Biopharm* **1**:175–200 (1973).

6. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

7. G. Levy, Mechanism-based pharmacodynamic modeling. *Clin Pharmacol Ther* **56**:356–358 (1994).

8. V. Garg and W. J. Jusko, Pharmacodynamic modeling of nonsteroidal anti-inflammatory drugs: antipyretic effect of ibuprofen. *Clin Pharmacol Ther* **55**:87–88 (1994).

9. W. J. Jusko and H. C. Ko, Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clin Pharmacol Ther* **56**:406–419 (1994).

10. D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz, Rapid turnover of plasma virion and CD4 lymphocytes in HIV-1 infection. *Nature* **373**:123–126 (1995).

11. X. Wei, S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, M. S. Saag, and G. S. Shaw, Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**:117–122 (1995).

12. A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**:1582–1586 (1996).

13. M. A. Nowak, S. Bonhoeffer, A. M. Hill, R. Boehme, H. C. Thomas, and H. McDade, Viral dynamics in hepatitis B virus infection. *Proc Natl Acad Sci USA* **93**:4398–4402 (1996).

14. S. Zeuzem, R. A. de Man, P. Honkoop, W. K. Roth, S. W. Schalm, and J. M. Schmidt, Dynamics of hepatitis B virus infection in vivo. *J Hepatol* **27**:431–436 (1997).

15. S. Zeuzem, J. M. Schmidt, J. H. Lee, B. Ruster, and W. K. Roth, Effect of interferon alfa on the dynamics of hepatitis C virus turnover in vivo. *Hepatology* **23**:366–371 (1996).

16. N. P. Lam, A. U. Neumann, D. R. Gretch, T. E. Wiley, A. S. Perelson, and T. J. Layden, Dose-dependent acute clearance of hepatitis C genotype 1 virus with interferon alfa. *Hepatology* **26**:226–231 (1997).

17. K. A. Powers, N. M. Dixit, R. M. Ribeiro, P. Golia, A. H. Talal, and A. S. Perelson, Modeling viral and drug kinetics: hepatitis C virus treatment with pegylated interferon alfa-2b. *Semin Liver Dis* **23**:13–18 (2003).

18. M. C. Rosario, P. Jacqmin, P. Dorr, E. van der Ryst, and C. Hitchcock, A pharmacokinetic–pharmacodynamic disease model to predict in vivo antiviral activity of maraviroc. *Clin Pharmacol Ther* **78**:508–519 (2005).

19. V. Garg, J. C. Ermer, J. Korth-Bradley, and S. Chiang, Comparison of effect-compartment and indirect response models for the analgesic effect of bromfenac in humans. *Pharm Res* **13**:S446 (1996).

20. I. F. Troco'niz, S. Armenteros, M. V. Planelles, J. Benitez, R. Calvo, and R. Dominguez, Pharmacokinetic–pharmacodynamic modeling of the antipyretic effect of two formulations of ibuprofen. *Clin Pharmacokinet* **38**:505–518 (2000).

21. J. M. T. Van-Griensven, W. J. Jusko, H. H. P. J. Lemkes, R. Kroon, C. J. Verhorst, S. T. Chiang, and A. F. Cohen, Tolrestat pharmacokinetic and pharmacodynamic effects on red blood cell sorbitol levels in normal volunteers and in patients with insulin-dependent diabetes. *Clin Pharmacol Ther* **58**:631–640 (1995).

22. K. H. Lew and W. J. Jusko, Pharmacodynamic modeling of cortisol suppression from fluocortolone. *Eur J Clin Pharmacol* **45**:581–583 (1993).

23. N. V. Nagaraja, B. Pechstein, K. Erb, C. Klipping, R. Hermann, M. Locher, and H. Derendorf, Pharmacokinetic/pharmacodynamic modeling of luteinizing hormone [LH] suppression and LH surge delay by cetrorelix after single and multiple doses in healthy premenopausal women. *J Clin Pharmacol* **43**:243–251 (2003).

24. F. R. Luo, Z. Yang, H. Dong, A. Camuso, K. McGlinchey, K. Fager, C. Flefleh, D. Kan, I. Inigo, S. Castaneda, T. W. Wong, R. A. Kramer, R. Wild, and F. Y. Lee, Prediction of active drug plasma concentrations achieved in cancer patients by pharmacodynamic biomarkers identified from the geo human colon carcinoma xenograft model. *Clin Cancer Res* **11**:5558–5565 (2005).

25. Y. Matsumoto, T. Fujita, Y. Ishida, M. Shimizu, H. Kakuo, K. Yamashita, M. Majima, and Y. Kumagai, Population pharmacokinetic–pharmacodynamic modeling of TF-505 using extension of indirect response model by incorporating a circadian rhythm in healthy volunteers. *Biol Pharm Bull* **28**:1455–1461 (2005).

26. M. H. Magee, R. A. Blum, C. D. Lates, and W. J. Jusko, Pharmacokinetic/pharmacodynamic model for prednisolone inhibition of whole blood lymphocyte proliferation. *J Clin Pharmacol* **53**:474–484 (2002).

27. R. A. A. Mathôt and W. P. Geus, Pharmacodynamic modeling of the acid inhibitory effect of ranitidine in patients in an intensive care unit during prolonged dosing: characterization of tolerance. *Clin Pharmacol Ther* **66**:140–151 (1999).

28. S. Choudhury, E. Nelson, K. Knub, and D. Hussain, Simultaneous pharmacokinetic and pharmacodynamic analysis to evaluate the induction response of tryptophan pyrrolase by hydrocortisone and tryptophan in rats. *Pharm Res* **11**:S411 (1994).

29. K. A. Nieforth, R. Nadeau, and I. H. Patel, Use of an indirect pharmacodynamic stimulation model of MX protein induction to compare in vivo activity of interferon alfa-2a and a polyethylene glycol-modified derivative in healthy subjects. *Clin Pharmacol Ther* **59**:636–646 (1996).

30. G. Movin-Osswald and M. Hammarlund-Udenaes, Prolactin release after remoxipride by an integrated pharmacokinetic–pharmacodynamic model with intra- and interindividual aspects. *J Pharmacol Exp Ther* **274**:921–927 (1995).

31. S. Choudhury, M. D. Hussain, and S. Vadlamudi, Pharmacokinetic–pharmacodynamic modeling of effect of haloperidol on plasma prolactin in rats. *Pharm Res* **11**:S424 (1994).

32. J. V. S. Gobburu, H. Agersø, W. J. Jusko, and L. Ynddal, Pharmacokinetic–pharmacodynamic modeling of ipamorelin, a growth hormone releasing peptide, in human volunteers. *Pharm Res* **16**:1412–1416 (1999).

33. C. A. Sarkar and D. A. Lauffenburger, Cell-level pharmacokinetic model of granulocyte colony-stimulating factor: implications for ligand lifetime and potency in vivo. *Mol Pharmacol* **63**:147–158 (2003).

34. N. A. Lack, B. Green, D. C. Dale, G. B. Calandra, H. Lee, R. T. MacFaarland, K. Badel, W. C. Liles, and G. Bridger, A pharmacokinetic–pharmacodynamic model for the mobilization of CD34$^+$ hematopoietic progenitor cells by AMD3100. *Clin Pharmacol Ther* **77**:427–436 (2005).

35. J. J. Lima, N. Matsushima, N. Kissoon, J. Wang, J. E. Sylvester, and W. J. Jusko, Modeling the metabolic effects of terbutaline in β_2-adrenergic receptor diplotypes. *Clin Pharmacol Ther* **76**:27–37 (2004).

36. J. M. Giraudel, A. Diquelou, V. Laroute, P. Lees, and P. L. Toutain, Pharmacokinetic/pharmacodynamic modeling of NSAIDs in a model of reversible inflammation in the cat. *Br J Pharmacol* **146**:642–653 (2005).

37. A. Sharma and W. J. Jusko, Characterization of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **24**:611–635 (1996).

38. A. Sharma and W. J. Jusko, Characteristics of indirect pharmacodynamic models and applications to clinical drug responses. *Br J Clin Pharmacol* **45**:229–239 (1998).

39. W. Krzyzanski and W. J. Jusko, Mathematical formalism and characteristics of four basic models of indirect pharmacodynamic responses for drug infusions. *J Pharmacokinet Biopharm* **26**:385–408 (1998).

40. L. B. Sheiner, D. R. Stanski, S. Vozeh, R. D. Miller, and J. Ham, Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to *d*-tubocurarine. *Clin Pharmacol Ther* **25**:358–371 (1979).

41. W. Krzyzanski, R. Ramakrishnan, and W. J. Jusko, Basic pharmacodynamic models for agents that alter production of natural cells. *J Pharmacokinet Biopharm* **27**:467–489 (1999).

42. G. J. Fetterly, J. M. Tamburlin, and R. M. Straubinger, Paclitaxel pharmacodynamics: application of mechanism-based neutropenia model. *Biopharm Drug Dispos* **22**:251–261 (2001).

43. W. Krzyzanski and W. J. Jusko, Multiple-pool cell lifespan model of hematologic effects of anticancer agents. *J Pharmacokinet Biopharm* **29**:311–337 (2002).

44. K. P. Zuideveld, H. J. Mass, N. Trejitel, J. Hushoff, P. H Van der Graff, L. A. Peletier, and M. Danhof, A set-point model with oscillatory behavior predicts the time-course of 8-OH-DPAT-induced hypothermia. *Am J Physiol* **281**:R2059–R2071 (2001).

45. K. P. Zuideveld, A. V. Gesel, L. A. Peletier, P. H. Van der Graff, and M. Danhof, Pharmacokinetic–pharmacodynamic modeling of the hypothermic and corticosterone effects of the 5-HT$_{1A}$ receptor agonist flesinoxan. *Eur J Pharmacol* **445**:43–54 (2002).

46. A. A. Fasanmade and W. J. Jusko, An improved pharmacodynamic model for formation of methemoglobin by antimalarial drugs. *Drug Metab Dispos* **23**:573–576 (1995).

47. W. C. Zamboni, D. Z. D'Argenio, C. F. Stewart, T. MacVittie, B. J. DeLauter, A. M. Farese, D. M. Potter, N. M. Dubat, D. Tubergen, and M. J. Egorin, Pharmacodynamic model of topotecan-induced time course of neutropenia. *Clin Cancer Res* **7**:2301–2308 (2001).

48. W. Krzyzanski and W. J. Jusko, Indirect pharmacodynamic models for response with multicompartment distribution or polyexponential disposition. *J Pharmacokinet Biopharm* **28**:57–78 (2002).

49. H. Li, G. M. Meno-Tetang, K. Chiba, N. Arima, P. Heining, and W. J. Jusko, Pharmacokinetics and cell trafficking dynamics of 2-amino-2[2-(4-octylphenyl)ethyl]propane-1,3-diol hydrochloride (FTY720) in cynomolgus monkeys after single oral and intravenous doses. *J Pharmacol Exp Ther* **301**:519–526 (2002).

50. W. Krzyzanski and W. J. Jusko, Mathematical formalism for the properties of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **25**:107–123 (1997).

51. W. Krzyzanski and W. J. Jusko, Mathematical formalism and characteristics of four basic models of indirect pharmacodynamics responses for drug infusions. *J Pharmacokinet Biopharm* **26**:385–408 (1998b).

52. E. A. Schaick, H. J. M. M. de Greef, A. P. Ijzerman, and M. Danhof, Physiological indirect effect modeling of the antilipolytic effects of adenosine A_1-receptor agonist. *J Pharmacokinet Biopoharm* **25**:673–694 (1997).

53. W. Krzyzanski and W. J. Jusko, Note: Caution in the use of empirical equations for pharmacodynamic indirect response models. *J Pharmacokinet Biopharm* **26**:735–741 (1998).

54. A. N. Kong, E. Ludwig, R. L. Slaughter, P. M. Distefano, J. Demasi, J. E. Middleton, and W. J. Jusko, Pharmacokinetics and pharmacodynamic modeling of direct suppression effects of methylprednisolone on serum cortisol and blood histamine in human subjects. *Clin Pharmacol Ther* **46**:616–628 (1989).

55. A. Chakraborty, W. Krzyzanski, and W. J. Jusko, Mathematical modeling of circadian cortisol concentrations using indirect response models: comparison of several methods. *J Pharmacokinet Biopharm* **27**:23–43 (1999).

56. W. Krzyzanski, A. Chakraborty, and W. J. Jusko, Algorithm for application of Fourier analysis for biorhythmic baselines of pharmacodynamic indirect response models. *Chronobiol Int* **17**:77–93 (2000).

57. J. A. Wald and W. J. Jusko, Corticosteroid pharmacodynamic modeling: osteocalcin suppression by prednisolone. *Pharm Res* **9**:1096–1098 (1992).

58. J. M. Pawlotsky, H. Dahari, A. U. Neumann, C. Hezode, G. Germanidis, I. Lonjon, L. Castera, and D. Dhumeaux, Antiviral action of ribavirin in chronic hepatitis C. *Gastroenterology* **126**:703–714 (2004).

59. M. Tsiang, J. F. Rooney, J. J. Toole, and C. S. Gibbs, Biphasic clearance kinetics of hepatitis B virus from patients during adefovir dipivoxil therapy. *Hepatology* **29**:1863–1869 (1999).

60. S. R. Lewin, R. M. Ribeiro, T. Walters, G. K. Lau, S. Bowden, S. Locarnini, and A. S. Perelson, Analysis of hepatitis B viral load decline under potent therapy: complex decay profiles observed. *Hepatology* **34**:1012–1018 (2001).

61. A. S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, and D. D. Ho, Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**:188–191 (1997).

62. A. S. Perelson, Modelling viral and immune system dynamics. *Nature* **2**:28–36 (2001).

63. A. A. Ding and H. Wu, Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* **2**:13–29 (2001).

64. H. Wu and A. A. Ding, Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**:410–418 (1999).

APPENDIX 22.1   IMPLEMENTATION OF MODELS IN WinNonlin/NONMEM

Indirect Response Model: Inhibition of Production Rate

WinNonlin Code

```
COMMANDS
NFUNCTIONS 1
NDERIVATIVES 1
NPARAMETERS 7
NCONSTANTS 1
PNAMES KA, F, dose, KE, V, IC50, KOUT; 'kₐ' 'F' 'dose' 'kₑ' 'V'
'IC₅₀' 'Kₒᵤₜ'
NSECONDARY 0
END
TEMP
RIN = CON(1); Initial values of response
END
#remark - define differential equations starting values
START
Z(1) = RIN
T=0
END
#remark - define differential equations
DIFFERENTIAL
CP = (KA*F*dose)*(exp(-KA*X)-exp(-KE*X))/((KE-KA)*V)  ; PK profile
after 1ˢᵗ order absorption

KIN = KOUT*RIN; at steady state
K0 = KIN*(1-(CP/(IC₅₀+CP))); KIN = zero order production rate
DZ(1) = K0 - KOUT*Z(1)
END
FUNCTION 1
F= Z(1)
END
#remark - define end of model
EOM
```

NONMEM Code

```
$PROB PK/PD INDIRECT RESPONSE MODEL (EXP1:INHIBITION OF KIN)
$DATA PKPD1.csv IGNORE=C
```

```
$INPUT TIME AMT DV ID CMT CLIN VIN
$SUBROUTINE ADVAN = 8 TOL =3
$MODEL NCOMP = 2
COMP=(CENTRAL)
COMP=(RESPONSE, DEFOBS)
$PK
CL=CLIN
V=VIN
KEL=CL/V
KIN=THETA(1)*EXP(ETA(1))   ;Zero order production rate
KOUT=THETA(2)*EXP(ETA(2))  ; First order elimination rate constant
of response
IC50=THETA(3)*EXP(ETA(3))
S1=V
F2=KIN/KOUT  ;Initial values of response (F2 = RIN=R₀)
$DES
DC1=A(1)/V
DADT(1)= -KEL*DC1
K0=KIN*(1-DC1/(IC50+DC1))
DADT(2)= K0-KOUT*A(2)
$ERROR
IPRE=F
Y = F*(1+ERR(1))+ERR(2)
$THETA
(0.1, 10)
(0.01, 0.1)
(0.001, 0.01)
$OMEGA
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
$SIGMA
0.1
5000
$EST MAXEVAL = 9999 SIG=6 METHOD =0 NOABORT
$SIM ONLYSIM (600000) SUBPROB=1$TABLE ID TIME DV IPRE
FILE=PKPD1.OUT
```

Life Span Model: Stimulatory Drug Effect and Zero-Order Input Model

WinNonlin Code

```
COMMANDS
NFUNCTIONS 1
NDERIVATIVES 1
NPARAMETERS 5
NCONSTANTS 2
```

```
PNAMES dose, KE, V, SC50, SMAX 'dose' 'ke' 'V' 'SC50''SMAX'
NSECONDARY 0

END
TEMP
RIN = CON(1); RIN =R0(Initial response)
TR = CON(2); TR = Life span
END
#remark - define differential equations starting values
START
Z(1) = RIN
T=0
END
#remark - define differential equations
DIFFERENTIAL
If X <= TR THEN
C1 = (dose/V)*exp(-KE*X); PK profile after IV administration
KIN = RIN/TR; KIN = Zero order production rate
ST = (1+((SMAX*C1)/(SC50+C1)))
STR = (1+((SMAX*0)/(SC50+0)))
DZ(1) = KIN*ST - KIN*STR
ENDIF
IF X > TR THEN
C1 = (dose/V)*exp(-KE*X); X = TIME
C2 = (dose/V)*exp(-KE*(X-TR))
KIN = RIN/TR; at steady state
ST = (1+((SMAX*C1)/(SC50+C1)))
STR = (1+((SMAX*C2)/(SC50+C2)))
DZ(1) = KIN*ST - KIN*STR
ENDIF
END
FUNCTION 1
F=Z(1)
END
#remark - define end of model
EOM
```

NONMEM Code

```
$PROB PK/PD INDIRECT RESPONSE MODEL (EXP1:INHIBITION OF KIN)
$DATA LIFESPAN1.csv IGNORE=C
$INPUT TIME AMT DV ID CMT CLIN VIN
$SUBROUTINE ADVAN = 8 TOL =3
$MODEL NCOMP = 3
COMP=(CENTRAL)
COMP=(RESPONSE, DEFOBS)
COMP=(DCENTRAL)
$PK CALLFL = -2
```

```
CL=CLIN
V=VIN
KEL=CLIN/VIN
KIN=THETA(1);*EXP(ETA(1))
SMAX=THETA(2);*EXP(ETA(2)); Maximum effect of drug
SC50=THETA(3);*EXP(ETA(3))
ALAG3=THETA(4);*EXP(ETA(4)); TR=ALAG3
S1=V
F2=KIN*ALAG3; RIN=KIN*TR = Initial values of response
  $DES
DC1=A(1)/V; Concentration at CMT1
DC3=A(3)/V; Concentration at CMT3
DADT(1)= -KEL*A(1)
DADT(3)=-KEL*A(3)
ST1=1+SMAX*DC1/(SC50+DC1)
ST3=1+SMAX*DC3/(SC50+DC3)
DADT(2)= KIN*ST1-KIN*ST3
$ERROR
IPRE=F
Y = F*(1+ERR(1))+ERR(2)
$THETA
(0.1, 10)
(0.1, 2)
(0.001, 0.01)
(0.1, 10)
$OMEGA
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
$SIGMA
0.1
5000
$EST MAXEVAL = 9999 SIG=6 METHOD =0 NOABORT
;$SIM ONLYSIM (600000) SUBPROB=1
$TABLE ID TIME DV IPRE FILE=lifespan1.OUT
```

HCV Viral Dynamic Models

WinNonlin Code

```
COMMANDS
NFUNCTIONS 1
NDERIVATIVES 2
NPARAMETERS 3
NCONSTANTS 5
PNAMES delta, eps, c; ('δ' 'ε' 'c')
NSECONDARY 1
SNAMES I0; (Initial number of productively infected cells)
```

```
END
TEMP
beta    = CON(1); β
P0      = CON(2) ; P0
V0      = CON(3); V0
Td      = CON(4); Td
X0      = CON(5); X0 = Lag time
END
#remark - define differential equations starting values
START
Z(1) = c*V0/P0
Z(2) = V0
END
#remark - define differential equations
DIFFERENTIAL
T=(delta*c)/(beta*P0); Remaining constant during treatments (1)
DZ(1) = betaZ(2)- delta*Z(1)
DZ(2) = (1-eps)*P0*Z(1)-c*Z(2)
END
FUNCTION 1
F= Z(2)
END
#remark - define any secondary parameters
SECONDARY
I0=c*V0/P0
END
#remark – define end of model
EOM
```

NONMEM Code

```
$PROB HCV viral kinetic models
$DATA exm.csv IGNORE=C
$INPUT TIME DV ID CMT AMT
$SUBROUTINE ADVAN = 8 TOL =3
$MODEL NCOMP = 2
COMP=(INFECT)
COMP=(HCV, DEFOBS)
$PK
BETA= 0.00000008 ; second order rate constant of de novo infection
rate (IU per ml)⁻¹ per hrs
PROD= 10 ;production rate constant of virus (IU per mL per cell
per hrs)
DEATH= 0.0001 ;death rate constant of target cells (per hrs)
VIN= 5130000 ;viral load (IU per ml)
XIN= 2.5 ;delay time of effectiveness (hrs)
DELTA= THETA(1) + ETA(1) ;infected cell death rate constant (per hrs)
EFC= THETA(2) + ETA(2) ;effectiveness of drug
```

```
C= THETA(3) + ETA(3) ;free virus clearnace rate constant (per hrs)
TIN= (DELTA*C)/(BETA*PROD) ;initial amount of target cells
;define differential equation starting values
F1=C*VIN/PROD
F2=5130000
$DES
DADT(1)= BETA*TIN*A(2)-DELTA*A(1)
DADT(2)= (1-EFC)*PROD*A(1)-C*A(2)
$ERROR
IPRED=F
Y = F*(1+ERR(1))+ERR(2)
$THETA
(0.001, 0.02, 1)
(0.5,0.9, 1)
(0.3, 0.5, 0.9)
$OMEGA
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
0 FIX ;(0.1, 5000)
$SIGMA
0.1
5000
$EST MAXEVAL = 9999 SIG=6 METHOD =0 NOABORT
$TABLE ID TIME DELTA EFC C FILE=EXM.OUT
$SCAT PRED VS DV
$SCAT (RES WRES) VS PRED
$SCAT (RES WRES) VS TIME
$SCAT IPRED VS DV
```

Mechanistic Pharmacokinetic/ Pharmacodynamic Models II

DONALD E. MAGER and WILLIAM J. JUSKO

23.1 INTRODUCTION

The major challenge of contemporary mechanism-based pharmacokinetic/pharmacodynamic (PK/PD) modeling is to characterize the time course of drug disposition and effects while revealing the pharmacological properties of the drug and the primary rate-limiting steps in the biology of the system (1, 2). Despite the vast array of pharmacological mechanisms of action and physiological processes that produce and control responses to drugs, the basic tenets of PD models remain the concepts of capacity limitation and the natural turnover of biological substances or functions. Capacity limitation often results from the law of mass action and limited densities of pharmacological targets, which are reflected in the traditional Hill function or sigmoidal E_{max} or I_{max} model (3, 4):

$$E = \frac{E_{max}C^{\gamma}}{EC_{50}^{\gamma} + C^{\gamma}} \tag{23.1a}$$

$$E = \frac{I_{max}C^{\gamma}}{IC_{50}^{\gamma} + C^{\gamma}} \tag{23.1b}$$

where capacity (E_{max}, I_{max}) and sensitivity (EC_{50}, IC_{50}) parameters define the nonlinear relationship between drug effect (E) and concentration in plasma or at a biophase (C). In contrast to this explicit equation, natural and pathophysiological turnover processes can be described using a simple differential equation:

$$\frac{dR}{dt} = k_{in} - k_{out}R, \quad R(0) = R^0 \tag{23.2}$$

where the rate of change of a biological factor (R) is determined by a production rate (k_{in}) and a first-order removal rate constant (k_{out}), and R^0 is the initial value.

It should be noted that k_{in} might range from a constant zero-order rate constant to complex functions responsible for irregular biorhythmic baseline profiles. Appreciation of these fundamental principles derives from the fact that both drugs and diseases alter normal biological cascades responsible for controlling the homeostasis of physiological systems.

In this chapter, the integration of capacity limitation and biosignal turnover concepts is revealed in an overview of mechanistic PD models for irreversible effects, transduction processes, and tolerance and rebound phenomena. Pertinent equations are provided along with most signature profiles and salient model features. This information may be useful in the design and analysis of relevant PD studies, and the cited references should be consulted for more details on the application of models for specific drugs or drug classes.

23.2 IRREVERSIBLE PHARMACOLOGICAL EFFECTS

Several classes of drugs, including chemotherapeutic compounds (antimicrobial, antiviral, antiparasitic, and anticancer drugs) and enzyme inhibitors, may interact with cells and/or proteins by an irreversible mechanism of action. Typically, such antagonists are involved in covalent binding interactions with pharmacological targets, which promote cell killing and/or the inactivation of specific proteins or enzymes. The potency of irreversibly acting agents can be difficult to quantify and interpret owing to time-dependent inactivation and significant temporal delays between drug exposure and effects. Thus, quantitative analyses with appropriate mechanistic PD models are critical for characterizing and understanding the pharmacology of these agents.

23.2.1 Cell Proliferation with Irreversible Inactivation

The following differential equation can be used to represent general cell proliferation with phase-nonspecific cell killing:

$$\frac{dR}{dt} = g(R) - f(C)R, \quad R(0) = R^0 \tag{23.3}$$

where the response variable (R) is cell number (e.g., malignant cells, bacteria, parasites, or viral load). A schematic of a simple PD system is shown in Figure 23.1

FIGURE 23.1 Schematic of PD models of cell proliferation with irreversible cell inactivation. Plasma or biophase drug concentrations (C) interact with cells (R), which are proliferating according to a growth model ($g(R)$), in an irreversible bimolecular manner ($f(C)$).

(5), where the proliferation of cells in the absence of drug, $g(R)$, is governed by a first-order rate constant of cell growth (k_g):

$$g(R) = k_s R - k_{deg} R = k_g R \tag{23.4}$$

which is determined by the difference between first-order rates of natural proliferation (k_s) and degradation (k_{deg}). Cell killing may be depicted as a bimolecular interaction between the drug and cell receptor, such that

$$f(C) = kC \tag{23.5}$$

where k is a second-order rate constant, and hence Eq. (23.3) becomes

$$\frac{dR}{dt} = k_g R - kCR \tag{23.6}$$

which may be solved explicitly to yield (5)

$$R = R^0 e(k_g t) \cdot e\left(-k \cdot AUC_0^t\right) \tag{23.7}$$

where R^0 is the initial value, $R(0)$, and AUC_0^t is the area under the concentration–time curve from zero to time (t). Providing that complete cell loss occurs with sufficient drug exposure, Eq. (23.7) predicts a log-linear relationship between survival fraction (R/R^0) and the drug dose. The slope of this relationship is a function of the affinity of the target cell for the drug along with the drug dose and clearance. Also, Eq. (23.6) may be used to define a so-called minimum inhibitory concentration (MIC), where $dR/dt = 0$ and thus $MIC = k_g/k$. Experimental data showing survival fractions of chimera spleen and osteosarcoma cells following single doses of cyclophosphamide to mice were used to validate the original derivation and dose–time–response relationships described by Eqs. (23.6) and (23.7) (5).

Owing to capacity-limited drug–cell contacts, the bimolecular interaction in Eq. (23.5) can be modified, whereby the second-order rate constant k is replaced by a maximum value (K_{max}) and a sensitivity parameter (KC_{50}) reflecting the drug concentration producing 50% of K_{max} (6):

$$f(C) = \frac{K_{max} C}{KC_{50} + C} \tag{23.8}$$

This alteration in the function of drug effect was necessary to characterize the effects of various intraperitoneal dose levels of piperacillin on the killing and growth dynamics of *Pseudomonas aeruginosa* in neutropenic mice (6). Simulations of expected PD profiles are shown in Figure 23.2. For these and subsequent simulations, simple monoexponential drug disposition is assumed ($C = C^0 e^{-k_{el} \cdot t}$) and all parameter values are listed in the figure legends. Whereas cells would grow exponentially in the absence of drug, these profiles show biphasic survival curves with an initial phase of cell killing followed by a regrowth phase as drug concentrations decline well below the KC_{50} value.

FIGURE 23.2 Simulated response profiles (solid lines) for cell proliferation model with irreversible inactivation (Figure 23.1), where $g(R)$ and $f(C)$ are given by Eqs. (23.4) and (23.8). Plasma drug concentrations are shown (dashed lines) for increasing intravenous doses ranging from 10 to 10,000 units. Parameter values are $k_{el} = 0.3\,h^{-1}$, $R^0 = 10,000$ units, $k_g = 0.01\,h^{-1}$, $K_{max} = 0.1\,h^{-1}$, and $KC_{50} = 10$ units.

Alternative models of cell growth dynamics may be substituted for Eq. (23.4) and tested using standard model fitting criteria. For example, in vitro and in vivo cell populations rarely continue to grow exponentially as a result of spatial, nutritive, and other factors that may place an upper limit on cell density (R_{ss}). The logistic growth model is one function that limits exponential growth and is defined as (7)

$$g(R) = k_g R \left(1 - \frac{R}{R_{ss}} \right) \tag{23.9}$$

Thus, when $R \ll R_{ss}$, $g(R) \approx k_g R$ and as $R \to R_{ss}$, $g(R) \to 0$. Careful control experiments in the absence of drug should be conducted which will support the cell growth function selected for modeling. Development of resistant bacteria was handled by Campion et al. (8) by considering two pools of cells with differing growth rate constants and sensitivities to cytotoxicity. Conversion of sensitive to resistant cells was handled as a first-order process. These general systems of equations may be utilized as appropriate initial frameworks upon which additional system complexities may be integrated as will be further demonstrated.

23.2.2   Cell Proliferation with Cycle-Specific Inactivation

Some chemotherapeutic agents, such as vinca alkaloids, taxanes, and camptothecins, only exert their effects during specific phases of the cell cycle (9). This property can be characterized with a PD model where the total cell population is conceptualized as distributing between two groups representing sensitive (R_S) and insensitive (R_I) cells (10). The model is shown in Figure 23.3 and can be described using the following general equations:

FIGURE 23.3   Schematic of PD models of cell proliferation with irreversible cycle-specific inactivation. The model shown in Figure 23.1 is modified to include a quiescent or insensitive pool of cells (R_I).

$$\frac{dR_s}{dt} = g(R_S) - f(C)R_S - k_{SI}R_S + k_{IS}R_I \tag{23.10a}$$

$$\frac{dR_I}{dt} = k_{SI}R_S - k_{IS}R_I \tag{23.10b}$$

where the functions of cell proliferation and killing are operable only on the R_S population, and the interconversion between groups is controlled by first-order rate constants (k_{SI} and k_{IS}). The initial total cell density (R_T^0) is defined as $R_T^0 = R_S^0 + R_I^0$ and thus the initial conditions of Eqs. (23.10a) and (23.10b) are

$$R_S^0 = \frac{\alpha}{1 + \alpha} R_T^0 \tag{23.11a}$$

$$R_I^0 = \frac{1}{1 + \alpha} R_T^0 \tag{23.11b}$$

where, assuming a simple exponential growth model (Eq. (23.4)),

$$\alpha = (0.5/k_{SI})\left[k_{IS} + k_g - k_{SI} + \sqrt{(k_{IS} + k_g - k_{SI})^2 + 4k_{IS}k_{SI}} \right] \tag{23.12}$$

Simulations of the cycle-specific inactivation model, utilizing a capacity-limited cell killing function (Eq. (23.8)), are shown in Figure 23.4. Whereas the profile resulting from a relatively low dose resembles those from the phase-nonspecific model (Figure 23.2), increasing dose levels produce biexponential cell-killing curves followed by exponential growth as drug concentrations decline below the KC_{50} value. The general model of cycle-specific inactivation (Eqs. (23.10a) and (23.10b)) is versatile and can easily be extended to include other proliferation functions such as a logistic growth model (7).

All of the previous models and equations have been applied to describe growth and killing of populations of individual cells such as bacteria, cancer cells, viruses, and parasites. More difficult is quantifying tumor size—weight or volume—where

FIGURE 23.4 Simulated response profiles (solid lines) for cell proliferation model with irreversible cycle-specific inactivation (Figure 23.3). Functions $g(R)$ and $f(C)$ are defined by Eqs. (23.4) and (23.8). Parameter values are $k_{el} = 0.3\,\text{h}^{-1}$, $R_T^0 = 10,100$ units, $k_g = 0.15\,\text{h}^{-1}$, $k_{SI} = 0.001\,\text{h}^{-1}$, $k_{IS} = 0.1\,\text{h}^{-1}$, $K_{max} = 1.0\,\text{h}^{-1}$, and $KC_{50} = 1.0$ unit.

the mass is heterogeneous in nature. Simeoni and co-workers (11) have applied an innovative chemotherapy model that employs cycle specificity, two growth rates (first exponential then linear), bimolecular inactivation by drug (Eq. (23.5)), and a series of transit compartments reflecting the transition of cells toward death. Section 23.4.2 will describe the type of transduction model that was utilized.

23.2.3 Turnover Model of Enzymatic Inactivation

Mechanistic enzyme inhibition represents a major therapeutic modality. Several inhibitors have been introduced clinically and have contributed significantly to the treatment of diseases such as gastric ulcers (H^+, K^+-ATPase inhibitors) and cancer (aromatase and thymidilate synthase inhibitors). PD models for such drugs must account for endogenous production and degradation of the target enzyme (Figure 23.5) and often utilize a modified indirect response turnover model (12):

$$\frac{dR}{dt} = k_{in} - k_{out}R - f(C)R \tag{23.13}$$

where R represents enzyme concentration or function, k_{in} is a constant zero-order production rate constant, k_{out} represents a drug-independent first-order removal rate constant, and $f(C)$ is as previously defined (Eq. (23.3)). In the absence of drug ($C = 0$), Eq. (23.13) collapses to a simple turnover function (Eq. (23.2)), where the baseline condition is $dR/dt = 0$ and thus $R^0 = k_{in}/k_{out}$. For purposes of stationarity and to reduce the number of model parameters to be estimated, k_{in} is often expressed as the product of R^0 and k_{out}. Simulated responses of the turnover model with irreversible inactivation (Eqs. (23.5) and (23.13)) are depicted in Figure 23.6. Profiles

FIGURE 23.5 Schematic of PD turnover models with irreversible inactivation. The production and loss of the response variable (R) typically are assumed to reflect zero-order (k_{in}) and first-order (k_{out}) rate processes.

FIGURE 23.6 Simulated response profiles (solid lines) for a turnover model with irreversible inactivation (Figure 23.5), where $f(C)$ is defined by Eq. (23.5). Parameter values are $k_{el} = 0.3\,h^{-1}$, $R^0 = 100$ units, $k_{out} = 0.1\,h^{-1}$, $k = 0.01$ unit, and k_{in} is specified as the product of k_{out} and R^0.

show a decline from an initial value to a nadir, followed by a gradual return to the baseline value as the drug is removed from the system. The time to maximal effect appears dose dependent with a shift to earlier times with increasing dose levels.

The basic turnover-irreversible effect model has been applied to various systems including the antiplatelet effect of aspirin (13) and the inhibition of gastric acid secretion by pantoprazole (14), an irreversible H^+, K^+-ATPase antagonist, based on the assumption that these functional measures are directly proportional to enzyme concentrations. The format of the model is flexible as well and has been extended to integrate sources of various complexities and data signatures. For example, precursor compartment pools have been added to model omeprazole dynamics in dogs (15) and the kinetics of dihydrotestosterone following exposure to 5α-reductase inhibitors in rats (16) and humans (17). In the case of omeprazole, differences in the rates of recovery from short- and long-term drug exposure were well characterized with the addition of a precursor (P), which was described by the following differential equation:

$$\frac{dP}{dt} = k_{in} - (k_1 + k_{out})P + k_2 R \tag{23.14}$$

The rate of change of the response variable was defined as

$$\frac{dR}{dt} = k_1 P - k_2 R - (k_{out} + kC)R \tag{23.15}$$

where transfer between compartments is controlled by first-order rate constants (k_1 and k_2), and both the precursor and response dissipate at a similar first-order rate (k_{out}). The initial conditions of Eqs. (23.14) and (23.15) are expressed as

$$P^0 = \frac{k_{in}}{k_{out}} \cdot \frac{(k_{out} + k_2)}{(k_1 + k_2 + k_{out})} \tag{23.16a}$$

$$R^0 = \frac{k_{in}}{k_{out}} \cdot \frac{k_1}{(k_1 + k_2 + k_{out})} \tag{23.16b}$$

Whereas short-term drug exposure may result in rapid returns to baseline values, long-term drug exposure depletes both the response variable and the precursor pool, which requires significantly longer durations of time for the baseline response value to be achieved once drug is removed.

23.2.4 Reactive Drug Metabolites

Irreversible bimolecular interaction models may be used to characterize the formation and effects of reactive drug metabolites. Several unrelated classes of drugs may undergo metabolism to form reactive intermediates that, when not sufficiently detoxified, may bind with cells and proteins and potentially elicit toxic effects (18). As an example, some antimalarial drugs form reactive metabolites that interact with hemoglobin (Hb) to form methemoglobin (MetHb), thereby causing cyanosis or methemoglobinemia (18). Fasanmade and Jusko (19) developed a PD model for the formation and disposition of MetHb following exposure of an antimalarial compound to dogs. The pharmacokinetics of the parent compound (C_p) were modeled first and then fixed as a driving function for fitting the PD model to the time course of MetHb concentrations. The rate of change of the reactive metabolite (C_m) was described as

$$\frac{dC_m}{dt} = k_f C_p - k_{me} C_m \tag{23.17}$$

where k_f and k_{me} are first-order rate constants of formation and elimination. Actual concentrations of the metabolite were not measured and thus hypothetical values were inferred during the model fitting process. The interaction between C_m and Hb to form MetHb is contained within the following differential equation:

$$\frac{dMetHb}{dt} = k \cdot Hb \cdot C_m - k_h \cdot MetHb \tag{23.18}$$

where k is now a second-order formation rate constant (as opposed to an elimination term as in Eq. (23.6)) and k_h is a first-order rate constant of MetHb elimination.

This model well characterized the time course of MetHb after a single oral dose of an antimalarial compound in dogs, including the substantial temporal lag between drug exposure and the peak of MetHb concentrations, which occurred several days following drug administration. This type of modeling of the profile of an unmeasured reactive metabolite may be relevant to other toxicological agents.

23.3 NONLINEAR AND TIME-DEPENDENT TRANSDUCTION PROCESSES

Drugs produce their pharmacological effects, both reversible and irreversible, through a variety of complex stimulus–response mechanisms (20). The complex nature of biological cascades has recast the concept of secondary messengers into that of signaling networks (21) and reintroduced the methodology of a systems analysis approach (22). In this section, PD models are presented for cases of nonlinear transduction or stimulus transfer as well as instances when signaling events represent a rate-limiting step in the production or loss of a pharmacological response.

23.3.1 Operational Model of Agonism

According to classical receptor occupancy and the law of mass action, the concentration of the drug–receptor complex (RC) for simple receptor binding is given by (3)

$$RC = R_T \cdot C/(K_D + C) \tag{23.19}$$

where R_T represents total receptor density and K_D is the equilibrium dissociation constant. The pharmacological effect is correlated to the bound receptor through a transducer function, z, such that

$$E = z(RC) \tag{23.20}$$

Assuming a linear transducer function, or that the effect is directly proportional to the drug–receptor complex ($E = \alpha \cdot RC$) (3), Eq. (23.20) becomes the empirical E_{max} model (Eq. (23.1); $\gamma = 1$), where E_{max} and EC_{50} replace the αR_T and K_D terms.

The Black–Leff (23) operational model of agonism offers a more mechanistic interpretation of concentration–effect curves and assumes a nonlinear transducer function:

$$E = \frac{E_m \cdot RC}{K_E + RC} \tag{23.21}$$

where E_m represents a system maximum and K_E is the RC concentration that elicits half-maximal effect. Thus, combining Eqs. (23.19) and (23.21) yields

$$E = \frac{E_m \tau C}{K_D + (\tau + 1)C} \tag{23.22}$$

where τ is defined as a measure of transduction efficiency ($\tau = R_T/K_E$). This equation results from one hyperbolic function (Eq. (23.19)) feeding into another (Eq. (23.21)) and is itself hyperbolic (20, 23). In this system, the transducer ratio (τ) influences both the maximum effect achievable by an agonist and the agonist concentration that produces 50% of the maximal effect. These two parameters are given by the following relationships:

$$E_{\max} = \frac{E_m \tau}{\tau + 1} \tag{23.23}$$

$$EC_{50} = \frac{K_D}{1 + \tau} \tag{23.24}$$

A slope term (n), analogous to the Hill coefficient (γ in Eq. (23.1)), also may be added to the transducer function (Eq. (23.21)), which transforms the concentration–effect relationship (Eq. (23.22)) to

$$E = \frac{E_m \tau^n C^n}{(K_D + C)^n + \tau^n C^n} \tag{23.25}$$

where in this case, the $E_{\max}$ and EC_{50} parameters are redefined as

$$E_{\max} = \frac{E_m \tau^n}{\tau^n + 1} \tag{23.26}$$

$$EC_{50} = \frac{K_D}{\left(2 + \tau^n\right)^{1/n} - 1} \tag{23.27}$$

Thus, the operational model of agonism resolves hyperbolic and sigmoidal concentration–effect curves utilizing drug-specific (K_D and τ) and system-specific (E_m) parameters. When available, directly measured receptor binding (Eq. (23.19)) may be integrated into the full relationship (Eq. (23.25)). The implementation of the model typically requires a comparative method where the pharmacodynamics of a series of compounds is examined simultaneously, or at least with prior knowledge of the properties of a full agonist. For example, estimates of E_m and n may be obtained from fitting the sigmoidal $E_{\max}$ model (Eq. (23.1)) to concentration–effect data obtained from full agonists, as $E_m = E_{\max}$ and $n = \gamma$ for such compounds. These terms are constrained in Eq. (23.25), which can then be used to estimate the drug-specific properties, namely, drug affinity and intrinsic efficacy, for partial agonists from fitting the equation to suitable concentration–effect profiles.

The effect of liposomal methylprednisolone on the inhibition of rat splenocyte proliferation provides an explicit example of Black–Leff principles in pharmacodynamics (24). Direct measurements of the drug–receptor complex were shown to account for the percent inhibition of lymphocyte proliferation for liposomal and free drug formulations jointly as specified by Eq. (23.21) (including a slope coefficient, n). The time course of in vivo lymphocyte proliferation was well captured by combining this relationship with suitable PK driving functions.

Van der Graaf and colleagues (25) applied the operational model in an integrated PK/PD analysis to characterize the in vivo pharmacodynamics of a series of adenosine A_1 receptor agonists in rats. Of importance was the observation that the in vivo estimates of K_D and τ were well correlated with direct experimental values obtained from in vitro bioassays. Therefore, assuming the system has been well described, the potential exists for predicting the time course of in vivo effects of relevant drug candidates from in vitro measurements.

The direct application of Eq. (23.25) to in vivo PD data assumes that drug concentrations in plasma and the biophase are in rapid equilibrium and directly proportional. Furthermore, maximum or peak effects are assumed to occur at peak drug concentrations (i.e., lack of hysteresis in concentration–effect curves). However, the operational model may be included also in indirect response models that characterize the temporal displacement between concentration and effect in mechanistic terms (26).

23.3.2 Transit Compartment Models for Signal Transduction Processes

Signal transduction cascades or networks, which are composed of a large array of secondary mediators, may represent a rate-limiting process and produce significant delays in pharmacological effects following drug administration. However, many of the individual steps involved in specific biosignaling cascades may be unknown or not readily measurable for many in vivo systems. Whereas empirical time-lag functions may be used to delay drug concentrations and thus serve as delayed driving functions in PD models, such an approach does not reflect the nature of the drug response and rarely captures the gradual onset of effect often observed for such systems. In contrast, the transit-compartment model (Figure 23.7) has been suggested as a robust platform for capturing time-dependent transduction delays in a mechanistic manner (27). Assuming reversible drug–receptor binding and a constant level of total receptors, the rate of change of the drug–receptor complex may be described by

$$\frac{dRC}{dt} = k_{on}(R_T - RC) \cdot C - k_{off} RC \qquad (23.28)$$

where k_{on} and k_{off} are binding microconstants representing rates of association (second-order) and dissociation (first-order). Subsequent signal transduction is characterized by a series of differential equations:

$$\frac{dM_1}{dt} = \frac{(RC - M_1)}{\tau} \cdots \frac{dM_i}{dt} = \frac{(M_{i-1} - M_i)}{\tau} \qquad (23.29)$$

FIGURE 23.7   Schematic of a time-dependent transduction model with three transit compartments (M_i) characterized by a mean transit time (τ). The production of the drug–receptor complex (RC) initiates the PD cascade and a linear transducer function (E^* and Eq. (23.20)) may be substituted for RC in the absence of specific receptor dynamics.

where M_i are the ith secondary messengers and τ represents a mean transit time. The pharmacological effect typically is expressed as $E = E^0 \pm M_N$, where N is the number of the last compartment in the series. In addition, a power coefficient (γ) may be incorporated into the series such that

$$\frac{dM_N}{dt} = \frac{(M_{N-1}^\gamma - M_N)}{\tau} \tag{23.30}$$

where γ can serve to amplify or dampen the response.

For most in vivo PD systems, where the receptor dynamics in Eq. (23.28) are unknown, a linear transducer function may be assumed (see Eq. (23.20)) and the sigmoidal E_{max} equation (Eq. (23.1)) may be substituted for RC in Eq. (23.29), thereby deriving a general PD model for time-dependent transduction processes (28). Interestingly, the simplest case, where $N = 1$, is mathematically equivalent to indirect response model I for inhibition ($E = E^0 - M_1$) and model III for stimulation ($E = E^0 + M_1$). Thus, the example in the previous section, where the operational model of agonism was included in an indirect response model, implies that a non-linear transducer function also may be used in Eq. (23.29), but this extension has yet to be evaluated with $N > 1$. Simulations of the general model, with $N = 3$, $\gamma = 1$, and $E^0 = 0$, are shown in Figure 23.8. Whereas peak drug concentrations are shown at $t = 0$, a delay and gradual onset of effect is observed, with maximal responses occurring at much later times in a dose-dependent manner. This signature profile is characteristic of PD delays owing to transduction processes. The parasympathomimetic activity of low-dose scopolamine and atropine in rats are examples for which this modeling approach has been applied (29).

Implementation of the general signal transduction model requires a search for an optimal number of transit compartments (N), which usually is the fewest that

FIGURE 23.8   Simulated response profiles (solid lines) for the signal transduction model shown in Figure 23.7. The E^* function is defined by Eq. (23.1) and the response is set equal to the time course of the M_3 compartment. Parameter values are $k_{el} = 0.3\,h^{-1}$, $E^0 = 0$ unit, $\tau = 5.0\,h$, $E_{max} = 100$ units, $\gamma = 1.0$, and $EC_{50} = 10$ units.

provide reasonable parameter estimates and model fitting. Careful attention to model fitting criteria (30), especially the distribution of residuals, must be made during model development. This necessitates a trial-and-error process; however, simulations using the gamma distribution function may provide initial guidance for further model refinement. The gamma distribution function is defined as

$$g_N(t) = \frac{k^N t^{N-1}}{(N-1)!} e^{-kt} \qquad (23.31)$$

where t is time and $k = 1/\tau$. The time course of the pharmacological effect may be approximated then by the product of the area under the effect curve and Eq. (23.31) (27).

An important feature of this modeling approach is its versatility, and many additional system complexities may be included in the final structure. As mentioned in Section 23.2.2, Simeoni and colleagues (11) combined an irreversible effect function with a transit-compartment model to describe tumor growth kinetics in nude mice xenograft models after the administration of several anticancer compounds. In contrast to expressing the response variable as a function of the last transit compartment, tumor weight was equated as the sum of all transit compartments reflecting a distribution of cells in various stages of cell death (see the appendix for example code). Clearly, time-dependent transduction models are easily applied and robust and may provide key insights and relevant predictions of drug effects using a minimal number of drug-specific and system-specific parameters.

23.4   TOLERANCE AND REBOUND PHENOMENA

Drug tolerance can be recognized as the diminishment or palliation of an expected PD response following repeated or continuous drug exposure. More complicated study designs must be employed to ensure repeated or lengthy drug administration and a sufficient washout period to capture the full return of the system to baseline conditions. Although processes involved in the development of tolerance often are complex and/or incompletely understood, PD models should attempt to characterize this property when present, as well as rebound or withdrawal phenomena, in mechanistic terms. The primary mechanisms responsible for tolerance and rebound include counterregulation, receptor desensitization, up- or downregulation of messenger RNA (mRNA), receptors, or secondary factors, and precursor pool alteration (Figure 23.9). Note that alternative PK sources of apparent drug tolerance, such as enzyme induction, antibody formation, and altered drug transport, have been excluded.

23.4.1   Counterregulatory Effect

PD models that take into account counterregulatory mechanisms rely on the generation of an opposing substance or effect. Although the structure of these models may take on various forms, the rate of change of an opposing mediator (M) typically is described by the following differential equation:

- **Counterregulation**

- **Desensitization**

- **Up– or Down– regulation**

- **Precursor Pool Alteration**

FIGURE 23.9 The primary mechanistic PD modeling approaches for functional adaptation or tolerance development.

$$\frac{dM}{dt} = k_1 R - k_2 M \qquad (23.32)$$

where the formation of M is governed by the product of the response variable (R) and a first-order rate constant (k_1), and k_2 is a first-order rate constant for the dissipation of the mediator. The values of k_1 and k_2 commonly are set equal to each other owing to difficulties associated with parameter identifiability. Regardless, the net PD response (R_{net}) may reflect the individual contributions of both R and M, where $R_{net} = R - M$. Bauer and Fung (31) developed a counterregulatory model to describe the hemodynamic effects and subsequent tolerance induced by a continuous infusion of nitroglycerin in a rat model of congestive heart failure. The response variable was considered to be directly proportional to plasma nitroglycerin (NTG) concentrations ($R = \alpha C_{NTG}$), whereas two transit compartments were used to generate the opposing biosignal and the net response was defined as $R_{net} = 100\% - R + M$.

In addition to simple additive effects, an opposing biosignal can be integrated into physiological PD models (30). Wakelkamp and colleagues (32) introduced a counterregulatory mechanism into an indirect response model of diuresis and later tolerance development following multiple intravenous doses of furosemide. The rate of change of the mediator was described by Eq. (23.32), with $k_{tol} = k_1 = k_2$. An indirect response model, where furosemide excretion rate (ER) was used to inhibit the first-order loss rate of the response variable, was modified accordingly:

$$\frac{dR}{dt} = k_{in} - k_{out}\left(1 - \frac{I_{max} \cdot ER}{IC_{50} + ER}\right) \cdot R \cdot (1 + M) \qquad (23.33)$$

Thus, an increase in the response variable due to the drug effect induces the generation of a mediator, which serves to stimulate the k_{out} term in direct opposition to the mechanism of action of the drug. Whereas the mediator in this example was hypothetical, it may represent an actual measurable substance. Lima and colleagues

(33) applied a feedback indirect response PK/PD model to characterize the effects of terbutaline on glucose–insulin homeostasis in healthy volunteers. Terbutaline plasma concentrations stimulate the production of glucose, which in turn stimulates the production and/or release of insulin. Insulin thus serves as the physiological mediator, where plasma concentrations above baseline conditions stimulate glucose efflux or utilization.

23.4.2 Receptor Desensitization

Another mechanism by which drug effects may lessen upon prolonged drug exposure is receptor desensitization, manifesting as receptor internalization or decreased apparent receptor affinity. The desensitization of G-protein-coupled receptors, which is mediated by protein kinases and exposure to select agonists, represents a classical example of this phenomenon (34). Although empirical functions may be used to alter traditional drug sensitivity parameters (e.g., K_D and EC_{50}) in a time- and/or exposure-dependent fashion, these models do not reflect the mechanisms involved and often fail to fully characterize this form of functional adaptation. In contrast, receptor inactivation theory encompasses both receptor occupancy and the rate theories of drug agonism and may embody a more mechanistic approach to receptor desensitization (20). The theory is described by the following series of equations:

$$\frac{dRC}{dt} = k_{on} \cdot R \cdot C - (k_{off} + k_3) \cdot RC \tag{23.34a}$$

$$\frac{dRC'}{dt} = k_3 \cdot RC - k_4 \cdot RC' \tag{23.34b}$$

$$\frac{dR}{dt} = -k_{on} \cdot R \cdot C + k_{off} \cdot RC + k_4 \cdot RC' \tag{23.34c}$$

where R stands for free receptor density, RC' is an inactive drug–receptor species, and k_3 and k_4 are first-order rate constants of RC' formation and loss. Whereas the rate of change of RC follows a traditional function (see Eq. (23.28)), its formation also drives the production of an inactive isoform (RC'). The drug effect in this system is assumed to be proportional to the rate of receptor inactivation (i.e., $E = \alpha k_3 \cdot RC$). Simulations of this model show that, given appropriate rate constants, transient peak responses occur followed by a gradual dose-dependent fade to steady-state values (20). However, this model has yet to be applied to in vivo PD systems.

23.4.3 Receptor or mRNA Up/Down-regulation

A third mechanism of tolerance is receptor/gene up- and down-regulation. The indirect response models are well suited for characterizing such effects (12), where drug exposure and/or the formation of a drug–receptor complex (RC) may serve to autoregulate the production or loss of the pharmacological target. Assuming that complete inhibition may be achieved, the rate of change of the target mRNA may be described as (35)

$$\frac{dmRNA}{dt} = k_{\text{in}}\left(1 - \frac{RC}{IC_{50} + RC}\right) - k_{\text{out}} \cdot mRNA \tag{23.35}$$

The ability of corticosteroids to inhibit the production of glucocorticoid receptor mRNA, and subsequently the formation of glucocorticoid receptor, is a classical example. This pharmacogenomic paradigm has been well characterized using the fifth-generation model for corticosteroid pharmacodynamics (35), capturing the depletion of glucocorticoid receptors and the development of tolerance following repetitive and continuous exposure of methylprednisolone to male adrenalectomized rats. The overall structure of this model has several important attributes that are discussed in Section 23.5.

23.4.4 Precursor Pool Alteration

A fourth major mechanism of tolerance and rebound is via either buildup or depletion of a precursor pool. Ariens (36) noted that certain drugs may cause a liberation of endogenous compounds that may require a significant amount of time to replenish once depleted. Therefore, if the endogenous substance is responsible for the desired pharmacological effect, then a form of tolerance may develop upon continued drug exposure. Precursor-dependent indirect response models have been evaluated for characterizing both tolerance and rebound phenomena (37). The rate of change of the precursor pool (P) and the response variable (R) can be described generally as

$$\frac{dP}{dt} = k_0 - k_p\{1 \pm H(C)\}P - k_s P \tag{23.36a}$$

$$\frac{dR}{dt} = k_p\{1 \pm H(C)\}P - k_{\text{out}}R \tag{23.36b}$$

where k_0 is a zero-order production rate of the precursor, k_p is a first-order rate constant of response production, k_s and k_{out} are first-order rate constants for the loss of the precursor and response, and $H(C)$ represents the I_{max} or E_{max} function (Eqs. (23.1a) and (23.1b); $\gamma = 1$). The need for an additional removal process for the precursor (k_s) may be tested using standard model fitting criteria. Plasma drug concentrations serve to stimulate ($+H(C)$) or inhibit ($-H(C)$) the processes involved in the production of the pharmacological response (k_p). The initial conditions for Eqs. (23.36a) and (23.36b) are

$$P^0 = k_0/(k_s + k_p) \tag{23.37a}$$

$$R^0 = k_p P^0 / k_{\text{out}} \tag{23.37b}$$

The model depicted at the bottom of Figure 23.9 reflects the two alteration of k_p options, while simulations of expected response profiles after escalating single-dose levels for the stimulation model are shown in Figure 23.10. Peak effects are shown to increase with dose, although not directly proportional, along with the time at which the peaks are achieved. Whereas the response variable for the lowest dose increases and then approaches the baseline value as drug concentrations decrease, increasing

FIGURE 23.10   Simulated response profiles (solid lines) for a precursor-dependent indirect response model (shown at the bottom of Figure 23.9). Drug exposure serves to stimulate the first-order rate conversion of a precursor to the response variable. Parameter values are $k_{el} = 0.3\,\text{h}^{-1}$, $R^0 = 15$ units, $P^0 = 300$ units, $k_p = 0.1\,\text{h}^{-1}$, $k_s = 0\,\text{h}^{-1}$, $k_{out} = 2.0\,\text{h}^{-1}$, $E_{max} = 1.0$, and $EC_{50} = 100$ units.

dose levels produce a rebound phenomenon where the response decreases below the baseline during the washout phase and then gradually returns toward the baseline. The extent of the rebound also is dose dependent. Additional simulations of responses under conditions of repeated drug administration or continuous infusion more clearly demonstrate the tolerance phenomenon predicted from this simple PD system (37). Movin-Osswald and Hammarlund-Udenaes (38) originally developed and applied this stimulatory precursor-dependent model ($k_s = 0$) to characterize the effect of remoxipride on prolactin release kinetics in healthy male volunteers. Although rebound was not pronounced in the response–time profiles, the data clearly showed tolerance development with consecutive intravenous administration, which was well described by the final model. Zannikos et al. (39) show data for drug effects on blockage of free fatty acid concentrations in plasma where both palliation of the inhibitory effect and a rebound occur. The data can be captured with the inhibitory precursor model, although the authors did not recognize this.

23.5   COMPLEX PHARMACODYNAMIC MODELS

Mechanistic modeling approaches to irreversible effects, transduction processes, and the development of tolerance have been presented, where the basic tenets of capacity limitation and physiological turnover processes were shown to be operational. As more primary determinants of drug action are determined through advances in molecular biology and pharmacology, so too are mechanism-based PD models likely to evolve and reflect integrated systems of basic modeling components. The

fifth-generation PK/PD model of corticosteroid pharmacogenomic effects (35) not only exemplifies this emerging paradigm, but its overall structure contains many of the major PD models discussed in this chapter. Thus, an overview of this model will serve to highlight and summarize fundamental techniques and provide insights into contemporary methods for complex PD systems analysis.

The full PK/PD model of acute corticosteroid receptor/gene-mediated effects is shown in Figure 23.11. PK functions will not be presented, but the model was primarily developed using intravenous and continuous infusion regimens of methylprednisolone to adrenalectomized rats, where the drug exhibits biexponential disposition and has been described using a two-compartment open model with linear first-order elimination from the central compartment. The rate of change of the drug–receptor complex is described as an irreversible process:

$$\frac{dRC}{dt} = k_{\text{on}} \cdot C \cdot R - k_t \cdot RC \tag{23.38}$$

Once formed, RC translocates into the cell nucleus ($RC(N)$) and this process is modeled using a transit compartment reflecting signal transduction:

$$\frac{dRC(N)}{dt} = k_t \cdot RC - k_{re} \cdot RC(N) \tag{23.39}$$

where k_t and k_{re} are first-order rate constants of the production and loss of the $RC(N)$ biosignal. A fraction of the activated receptors will return to the free pool of available glucocorticoid receptors (R_f in the model diagram). The activated drug–receptor complex ($RC(N)$) is the driving function controlling the corticosteroid pharmacological effects. As discussed previously, the formation of $RC(N)$ downregulates the synthesis of the mRNA for the glucocorticoid receptor (R) (see analogous Eq. (23.35)):

$$\frac{dmRNA_R}{dt} = k_{\text{syn\_Rm}} \cdot \left(1 - \frac{RC(N)}{IC_{50\_Rm} + RC(N)} \right) - k_{\text{deg\_Rm}} \cdot mRNA_R \tag{23.40}$$

FIGURE 23.11 The fifth-generation model of corticosteroid pharmacogenomic effects Symbols are defined in the text. (Adapted from Ref. 35.)

where k_{syn_Rm} and k_{deg_Rm} are zero- and first-order rate constants of mRNA production and degradation. The synthesis of free receptors is linked to the time course of its mRNA:

$$\frac{dR}{dt} = k_{syn\_R} \cdot mRNA_R + R_f \cdot k_{re} \cdot RC(N) - k_{on} \cdot C \cdot R - k_{dgr\_R} \cdot R \qquad (23.41)$$

where k_{syn_R} and k_{dgr_R} are first-order rate constants of receptor synthesis and degradation. Drug exposure will decrease the availability of free receptors as well as the rate at which the new receptors can be synthesized, thus imposing drug tolerance.

The induction of hepatic tyrosine aminotransferase (TAT) mRNA and activity is a classical measure of corticosteroid-mediated pharmacogenomic effects. The TAT dynamics are modeled in a manner consistent with precursor-dependent models, where the stimulation function acts on the production rate of the precursor species:

$$\frac{dmRNA_{TAT}}{dt} = k_{syn\_tm} \cdot \{1 + S \cdot RC(N)\} - k_{dgr\_tm} \cdot mRNA_{TAT} \qquad (23.42a)$$

$$\frac{dTAT}{dt} = EF \cdot (mRNA_{TAT})^{\gamma} - k_{dgr\_t} \cdot TAT \qquad (23.42b)$$

where k_{syn_tm} is a zero-order rate constant of TAT mRNA synthesis, S is a linear coefficient for the efficiency of TAT gene induction, k_{dgr_tm} is a first-order rate constant for TAT mRNA degradation, EF is the translational efficiency of TAT mRNA to the enzyme (or activity), γ is an amplification power coefficient, and k_{dgr_t} is a first-order degradation rate constant for TAT. The original article should be consulted for definitions of baselines and initial conditions.

The construction of such a complex model was achieved in a piecewise manner. For example, drug pharmacokinetics were measured, modeled, and fixed as the primary driving function. A single intravenous bolus dose of methylprednisolone produces a rapid drop in glucocorticoid receptor density, which gradually returns toward baseline values over 72 hours in a complex biphasic manner reflecting receptor recycling and de novo synthesis. The time course of glucocorticoid receptor mRNA shows a gradual decrease to a nadir at about 10 hours followed by a return to baseline within 48 hours. Measurements of receptor density and mRNA may be modeled simultaneously to characterize receptor dynamics and generate the activated drug–receptor biosignal. As with the PK function, the estimated model parameters may be fixed for subsequent fitting of the TAT mRNA and enzyme activity measurements. The shapes of these PD profiles resemble the simulations shown in Figure 23.8, where a lag-time and gradual onset is observed, with peak TAT mRNA density occurring around 5 hours. TAT activity dynamics appear similar but shifted slightly to the right, peaking around 7 hours, and both mRNA and TAT activity return toward baseline values by 18 hours. Complex models such as the one depicted in Figure 23.11 and the iterative manner in which it was constructed and experimentally validated may become commonplace as the focus of PK/PD modeling continues to shift toward integrative and systems pharmacology.

ACKNOWLEDGMENTS

REFERENCES

1. D. E. Mager, E. Wyska, and W. J. Jusko, Diversity of mechanism-based pharmacodynamic models. *Drug Metab Dispos* **31**:510–519 (2003).

2. G. Levy, Mechanism-based pharmacodynamic modeling. *Clin Pharmacol Ther* **56**:356–358 (1994).

3. E. J. Ariens, Affinity and intrinsic activity in the theory of competitive inhibition. *Arch Int Pharmacodyn Ther* **99**:32–49 (1954).

4. J. G. Wagner, Kinetics of pharmacological response: I. Proposed relationships between response and drug concentration in the intact animal and man. *J Theor Biol* **20**:173–201 (1968).

5. W. J. Jusko, Pharmacodynamics of chemotherapeutic effects: dose–time–response relationships for phase-nonspecific agents. *J Pharm Sci* **60**:892–895 (1971).

6. J. G. Zhi, C. H. Nightingale, and R. Quintiliani, Microbial pharmacodynamics of piperacillin in neutropenic mice of systematic infection due to *Pseudomonas aeruginosa*. *J Pharmacokinet Biopharm* **16**:355–375 (1988).

7. T. B. Robertson, *The Chemical Basis of Growth and Senescence*. Lippincott, Philadelphia, 1923.

8. J. J. Campion, P. J. McNamara, and M. E. Evans, Pharmacodynamic modeling of ciprofloxacin resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **49**:209–219 (2005).

9. B. A. Chabner, D. P. Ryan, L. Paz-Ares, R. Garcia-Carbonero, and P. Calabresi, in *The Pharmacological Basis of Therapeutics*, 10th ed., J. G. Hardman, L. E. Limbird, and A. G. Gilman (Eds.). McGraw-Hill, New York, 2001, pp. 1389–1459.

10. W. J. Jusko, A pharmacodynamic model for cell-cycle-specific chemotherapeutic agents. *J Pharmacokinet Biopharm* **1**:175–200 (1973).

11. M. Simeoni, P. Magni, C. Cammia, G. De Nicolao, V. Croci, E. Pesenti, M. Germani, I. Poggesi, and M. Rocchetti, Predictive pharmacokinetic–pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res* **64**:1094–1101 (2004).

12. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

13. K. Yamamoto, M. Abe, M. Katashima, Y. Yamada, Y. Sawada, and Y. Iga, Pharmacodynamic analysis of antiplatelet effect of aspirin in the literature—modeling cyclooxygenase in the platelet and the vessel wall endothelium. *Jpn J Hosp Pharm* **22**:133–141 (1996).

14. G. M. Ferron, W. McKeand, and P. R. Mayer, Pharmacodynamic modeling of pantoprazole's irreversible effect on gastric acid secretion in humans and rats. *J Clin Pharmacol* **41**:149–156 (2001).

15. A. Abelo, U. G. Eriksson, M. O. Karlsson, H. Larsson, and J. Gabrielsson, A turnover model of irreversible inhibition of gastric acid secretion by omeprazole in the dog. *J Pharmacol Exp Ther* **295**:662–669 (2000).

16. M. Katashima, K. Yamamoto, Y. Tokuma, T. Hata, Y. Sawada, and T. Iga, Pharmaco-kinetic and pharmacodynamic study of a new nonsteroidal 5 alpha-reductase inhibitor, 4-[3-[3-[bis(4-isobutylphenyl)methylamino]benzoyl]-1h-indol-1-Yl]-butyric acid in rats. *J Pharmacol Exp Ther* **284**:914–920 (1998).

17. P. O. Gisleskog, D. Hermann, M. Hammarlund-Udenaes, and M. O. Karlsson, A model for the turnover of dihydrotestosterone in the presence of the irreversible 5 alpha-reductase inhibitors Gi198745 and finasteride. *Clin Pharmacol Ther* **64**:636–647 (1998).

18. M. Pirmohamed and B. K. Park, in *Handbook of Drug Metabolism*, T.F. Woolf (Ed.). Marcel Dekker, New York, 1999, pp. 443–476.

19. A. A. Fasanmade and W. J. Jusko, An improved pharmacodynamic model for formation of methemoglobin by antimalarial drugs. *Drug Metab Dispos* **23**:573–576 (1995).

20. T. Kenakin, *Pharmacologic Analysis of Drug–Receptor Interactions*. Lippincott–Raven, Philadelphia, 1997.

21. U. S. Bhalla and R. Iyengar, Emergent properties of networks of biological signaling pathways. *Science* **283**:381–387 (1999).

22. H. Kitano, Systems biology: a brief overview. *Science* **295**:1662–1664 (2002).

23. J. W. Black and P. Leff, Operational models of pharmacological agonism. *Proc R Soc Lond B Biol Sci* **220**:141–162 (1983).

24. E. V. Mishina and W. J. Jusko, Inhibition of rat splenocyte proliferation with methylpred-nisolone: in vivo effect of liposomal formulation. *Pharm Res* **11**:848–854 (1994).

25. P. H. Van der Graaf, E. A. Van Schaick, R. A. Mathot, A. P. Ijzerman, and M. Danhof, Mechanism-based pharmacokinetic–pharmacodynamic modeling of the effects of N6-cyclopentyladenosine analogs on heart rate in rat: estimation of in vivo operational affinity and efficacy at adenosine A1 receptors. *J Pharmacol Exp Ther* **283**:809–816 (1997).

26. P. H. Van der Graaf, E. A. Van Schaick, S. A. Visser, H. J. De Greef, A. P. Ijzerman, and M. Danhof, Mechanism-based pharmacokinetic–pharmacodynamic modeling of antili-polytic effects of adenosine A1 receptor agonists in rats: prediction of tissue-dependent efficacy in vivo. *J Pharmacol Exp Ther* **290**:702–709 (1999).

27. Y. N. Sun and W. J. Jusko, Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J Pharm Sci* **87**:732–737 (1998).

28. D. E. Mager and W. J. Jusko, Pharmacodynamic modeling of time-dependent transduc-tion systems. *Clin Pharmacol Ther* **70**:210–216 (2001).

29. I. Perlstein, D. Stepensky, W. Krzyzanski, and A. Hoffman, A signal transduction phar-macodynamic model of the kinetics of the parasympathomimetic activity of low-dose scopolamine and atropine in rats. *J Pharm Sci* **91**:2500–2510 (2002).

30. J. Gabrielsson and D. Weiner, *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. Swedish Pharmaceutical Press, Stockholm, 1997.

31. J. A. Bauer and H. L. Fung, Pharmacodynamic models of nitroglycerin-induced hemo-dynamic tolerance in experimental heart failure. *Pharm Res* **11**:816–823 (1994).

32. M. Wakelkamp, G. Alvan, J. Gabrielsson, and G. Paintaud, Pharmacodynamic modeling of furosemide tolerance after multiple intravenous administration. *Clin Pharmacol Ther* **60**:75–88 (1996).

33. J. J. Lima, N. Matsushima, N. Nissoon, J. Wang, J. E. Sylvester, and W. J. Jusko, Modeling the metabolic effects of terbutaline in beta2-adrenergic receptor diplotypes. *Clin Pharmacol Ther* **76**:27–37 (2004).

34. J. C. Foreman and T. Johansen, *Textbook of Receptor Pharmacology*. CRC Press, Boca Raton, FL, 1996.

35. R. Ramakrishnan, D. C. DuBois, R. R. Almon, N. A. Pyszczynski, and W. J. Jusko, Fifth-generation model for corticosteroid pharmacodynamics: application to steady-state

receptor down-regulation and enzyme induction patterns during seven-day continuous infusion of methylprednisolone in rats. *J Pharmacokinet Pharmacodyn* **29**:1–24 (2002).

36. E. J. Ariens, *Molecular Pharmacology*. Academic Press, New York, 1964.

37. A. Sharma, W. F. Ebling, and W. J. Jusko, Precursor-dependent indirect pharmacodynamic response model for tolerance and rebound phenomena. *J Pharm Sci* **87**:1577–1584 (1998).

38. G. Movin-Osswald and M. Hammarlund-Udenaes, Prolactin release after remoxipride by an integrated pharmacokinetic–pharmacodynamic model with intra- and interindividual aspects. *J Pharmacol Exp Ther* **274**:921–927 (1995).

39. P. N. Zannikos, S. Rohatagi, and B. K. Jensen, Pharmacokinetic–pharmacodynamic modeling of the antipolytic effects of an adenosine receptor agonist in healthy volunteers. *J Clin Pharmacol* **41**:61–69 (2001).

APPENDIX 23.1

The following example illustrates the integration of irreversible effects and time-dependent transduction modeling to characterize in vivo tumor growth kinetics after exposure to anticancer drugs. This model was developed by Simeoni and colleagues[1] and was evaluated using several established and candidate chemotherapeutic agents. In this example, tumor weight was measured in HCT116 tumor-bearing mice (female Hsd, athymic nude-*nu* mice) after receiving a single intravenous injection of either the vehicle or irinotecan (CPT-11; doses were 45 or 60 mg/kg on day 13 following tumor transplantation). CPT-11 pharmacokinetics was determined in a separate analysis and characterized by an open two-compartment mamillary model with first-order elimination from the central compartment.

The model can be described by the following system of differential equations:

$$\frac{dM_1}{dt} = g(t) - k \cdot C \cdot M_1, \quad M_1(0) = w(0)$$

$$\frac{dM_2}{dt} = k \cdot C \cdot M_1 - M_2/\tau, \quad M_2(0) = 0$$

$$\frac{dM_3}{dt} = (M_2 - M_3)/\tau, \quad M_3(0) = 0$$

$$\frac{dM_4}{dt} = (M_3 - M_4)/\tau, \quad M_4(0) = 0$$

where M_i represents the ith transit compartment, $g(t)$ is the equation describing tumor growth, k is a second-order rate constant of drug effect, C represents CPT-11 plasma concentration, and τ is a mean transit time. Tumor weight, $w(t)$, is equated as the sum of the transit compartments: $w(t) = \sum_{i=1}^{4} M_i$, and tumor growth, $g(t)$, is defined by the following equation:

---

[1] Predictive pharmacokinetic–pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res* **64:**1094–1101 (2004).

$$g(t) = \frac{k_g \cdot M_1}{\left[1 + \left(\dfrac{k_g}{k_{g0}} \cdot w(t)\right)^{\Psi}\right]^{1/\Psi}}$$

where k_g and k_{g0} are first- and zero-order growth rate constants. According to the original derivation, large values of the Ψ parameter provide a good approximation for the system where tumor growth switches from a first-order to a zero-order rate. This term is fixed to 20 and the original article should be consulted for further details.

The above model was implemented in WinNonlin (Pharsight, Mountain View, CA) and the accompanying data were simulated using the previous parameter estimates ($k = 3.51 \times 10^{-5}\,\mathrm{ng^{-1}\,mL\cdot h^{-1}}$, $\tau = 51.2\,\mathrm{h}$, $k_g = 0.00608\,\mathrm{h^{-1}}$, $k_{g0} = 0.0139\,\mathrm{g/h}$, $w(0) = 0.085\,\mathrm{g}$) with $\pm 10\%$ random error.

```
Model 1
remark ****************************************************
remark Model: Irreversible Transduction Pharmacodynamic
remark Model of Tumor Growth
remark
remark Example: CPT-11 effects on tumor growth kinetics
remark in HCT116 tumor-bearing mice
remark
remark Developed by Simeoni et al. Cancer Res 2004 64:1094
remark Coded by Mager and Jusko 2005
remark ****************************************************
remark
remark - define model-specific commands
COMMANDS
NFUNCTIONS 3
NDERIVATIVES 9
NPARAMETERS 5
PNAMES 'tau', 'k', 'kg', 'kg0','w0'
END
remark - define temporary variables
TEMPORARY
T = X
remark - CPT-11 PK parameters
V = 4.85
k10 = 0.553
k12 = 0.0115
k21 = 0.0616
D1 = 45*1000
D2 = 60*1000

alpha = 0.5*((k12+k21+k10)+SQRT((k12+k21+k10)**2-4*k21*k10))
beta = 0.5*((k12+k21+k10)-SQRT((k12+k21+k10)**2-4*k21*k10))
C11 = D1*(alpha-k21)/(V*(alpha-beta))
```

```
C12 = D1*(k21-beta)/(V*(alpha-beta))
C21 = D2*(alpha-k21)/(V*(alpha-beta))
C22 = D2*(k21-beta)/(V*(alpha-beta))

END
remark - define differential equations starting values
START
Z(1) = w0
Z(2) = 0
Z(3) = 0
Z(4) = 0
Z(5) = w0
Z(6) = 0
Z(7) = 0
Z(8) = 0
Z(9) = w0
END
remark - define differential equations
DIFFERENTIAL

remark - IV bolus doses administered at t=312
IF T<312 THEN
Cp1=0
Cp2=0
ELSE
Cp1=C11*DEXP(-alpha*(t-312))+C12*DEXP(-beta*(t-312))
Cp2=C21*DEXP(-alpha*(t-312))+C22*DEXP(-beta*(t-312))
ENDIF

remark - Tumor growth kinetics
psi=20
wt1=Z(1)+Z(2)+Z(3)+Z(4)
wt2=Z(5)+Z(6)+Z(7)+Z(8)
wt3=Z(9)

gr1=kg*Z(1)/((1+(kg/kg0*wt1)**psi)**(1/psi))
gr2=kg*Z(5)/((1+(kg/kg0*wt2)**psi)**(1/psi))
gr3=kg*Z(9)/((1+(kg/kg0*wt3)**psi)**(1/psi))
remark - PD system equations
DZ(1) = gr1-k*Cp1*Z(1)
DZ(2) = k*Cp1*Z(1)-Z(2)/tau
DZ(3) = (Z(2)-Z(3))/tau
DZ(4) = (Z(3)-Z(4))/tau
DZ(5) = gr2-k*Cp2*Z(5)
DZ(6) = k*Cp2*Z(5)-Z(6)/tau
DZ(7) = (Z(6)-Z(7))/tau
DZ(8) = (Z(7)-Z(8))/tau
DZ(9) = gr3
```

```
END
remark - define algebraic functions
FUNCTION 1
F= Z(1)+Z(2)+Z(3)+Z(4)
END
FUNCTION 2
F= Z(5)+Z(6)+Z(7)+Z(8)
END
FUNCTION 3
F= Z(9)
END
remark - end of model
EOM
```

PK/PD Analysis of Binary (Logistic) Outcome Data

JILL FIEDLER-KELLY

24.1 INTRODUCTION AND APPLICATION OF LOGISTIC REGRESSION IN NONLINEAR MIXED EFFECTS MODELING

Binary outcome data, or endpoints with exactly two possible outcomes, are commonly collected during drug development. Examples of binary (also referred to as dichotomous) endpoints include cure versus lack of cure of a disease or condition with treatment, relief versus lack of relief from symptoms, eradication versus persistence of an organism, presence versus absence of a medical outcome, and appearance versus lack of appearance of an adverse event. Binary data are a subset of what are termed discrete or categorical endpoint data. Sheiner and Beal (1) have referred to such endpoints as "odd-type" data, in that they are noncontinuous and therefore require the use of nonstandard methodology for proper analysis and interpretation.

While binary outcome data may seem relatively straightforward and amenable to interpretation, often the determination of such endpoints depends on the utilization and understanding of many different (discrete and continuous) variables and is a function of a number of different factors. In the analysis of anti-infective compounds, a common efficacy endpoint is "clinical cure," which may be a function of microbiological results, patient-reported symptoms, and investigator examination. Taken together, these data are utilized (sometimes with an equation or mathematical function) to establish a determination of clinical cure or clinical failure for a particular patient. Other applications include situations where the observed dichotomous response is a mere simplification of an underlying latent or observed continuous response. An example of a latent continuous response may be craving for a cigarette. Here, the underlying continuous scale, although subjective and prone to a considerable amount of between-subject variability, may be simplified by creating levels or categories of response (0 = no craving, 1 = mild craving, 2 = moderate craving, 3 = intense craving; or more simply, 0 = mild or no craving and 1 = moderate to severe craving). Oftentimes, binary endpoints are created by

categorizing observed continuous responses based on a clinically significant cut-point. For example, observed continuous laboratory values that are greater than 3 times the upper limit of normal may be classified as an endpoint to be studied in comparison to the group of all other possible lower values.

Traditional statistical analyses of dichotomous dependent variables (endpoints) usually involve logistic regression analysis. Logistic regression models, in particular, linear logistic regression models, are a special case of the general linear models (GLMs) (2), and they provide a general unifying framework for the analysis of binary data as general linear models do for normally distributed continuous data. Logistic regression models utilize the logistic transformation or logit (defined in Section 24.2) to deal with the "problem" inherent in the response classification of only 0 or 1. The logit transforms the 0 to 1 probability scale to a $-\infty$ to $+\infty$ scale, allowing for the development of linear models describing the relationship between the success probability and various predictors; similarly, the fitted probabilities from these models are bounded between 0 and 1.

In pharmacometric analyses, exposure–response models are often developed using a population pharmacokinetic/pharmacodynamic (PK/PD) approach, combining multiple continuous PK observations (drug concentrations), and single or multiple efficacy (cure versus fail) and/or safety endpoints (occurrence or lack of occurrence of an adverse event) as the dependent variables. There are many examples of the use of logistic regression techniques to model binary endpoints in population PK/PD analyses (3–6). Johnson et al. (3) utilized a population approach to estimate the pharmacokinetics of orally administered midazolam and its 1-hydroxy metabolite and then a more traditional logistic regression analysis to estimate the effects of plasma midazolam and 1-hydroxymidazolam on sedation score in children undergoing surgery. Johnston et al. (4) utilized NONMEM® (7) to estimate the effects of exposure and other covariates on the probability of efficacy and risk of certain adverse events associated with the use of bupropion SR in smoking cessation. Mould et al. (5) and Xie et al. (6) utilized population PK/PD methods to estimate the more complicated probability of graded adverse events (ordinal responses) associated with treatment and their relationship with drug exposure.

While population PK analyses often involve the utilization of mixed effects models due to the repeated nature of the measurements collected from each individual and the desire to estimate and discriminate between the various sources of variability, PK/PD analyses of binary endpoint data may utilize either fixed or mixed effects models. Oftentimes, a single endpoint measurement is collected from each individual being studied and a model estimating only fixed effects is used. However, when multiple observations are collected from each individual (over time), we may wish to estimate the change in response probability over time while recognizing the correlation between observations from the same individual and also estimate the variation between individuals. This is accomplished through the use of a mixed effects model.

There are many examples of PK/PD exposure–response analyses conducted subsequent to population PK model development using a traditional statistical package such as SAS® (8) either with or without the use of a mixed effects model (9–11). As Yano et al. (12) have demonstrated in their simulations, this approach may be perfectly adequate in many circumstances and the assumed gain in precision of the fixed effect estimates through the use of a mixed effects model is either marginal

or nonexistent. Importantly, however, while careful attention to study design and data collection makes the simultaneous evaluation and estimation of PK and PD endpoints possible, even with sparse data, Yano et al. (12) point out that such sophisticated methods are not always warranted or even advantageous. While all software packages allowing for the estimation of logistic regression models differ, for the pharmacometrician who is experienced with the use of NONMEM for PK analysis, the NONMEM software provides an excellent framework for the implementation of a range of simple to complex models.

The theory and techniques described in this chapter focus on the application of logistic regression to binary outcome data and the development of models to describe the relationship between binary endpoints and one or more explanatory variables (covariates). While many software options are available for fitting fixed or mixed effects logistic regression models, this chapter endeavors to illustrate the use of nonlinear mixed effects modeling to analyze binary endpoint data as implemented in the NONMEM software.

24.2 STATISTICAL BASIS FOR LOGISTIC REGRESSION MODELS

We may consider the response of a particular patient, administered a given treatment, as binary if it can be quantified as either success or failure. We denote this response by R, and it is valued 1 (indicating success) or 0 (indicating failure). The unknown probability of a successful response is denoted by $p = P(R = 1)$. The corresponding probability of failure is denoted by $P(R = 0) = 1 - p$. The probability distribution of R can then be denoted by $P(R = r) = p^r(1 - p)^{1-r}$, $r = 0, 1$. This distribution is known as the *Bernoulli distribution* (13).

Building on this notation for a single patient, the individual binary responses of a group of n patients administered the given treatment can be thought of as a series of Bernoulli "trials" and described using the binomial distribution, with

$$P(Y = y) = (_nC_y)p^y(1 - p)^{n-y}, \quad y = 0, 1, \ldots, n$$

Here, y is the total number of patients with successful outcomes and we assume that each patient's response is independent of every other patient's response. Then, the probability of a particular distribution of successes and failures for a group of n patients is $p^y(1 - p)^{n-y}$ and the total number of ways in which a sequence of y successes and $n - y$ failures can occur is

$$(_nC_y) = \frac{n!}{y!(n - y)!}$$

Since we are most often concerned with developing a model that will describe the relationship between some binary endpoint (y) and one or more predictor variables (x), we are often interested in the conditional mean of y (the endpoint) given x, or $E(Y|x)$. In linear regression, we express this expectation as a linear function: $E(Y|x) = \beta_0 + \beta_1 x$. However, given the binary nature of our endpoint (y), this expectation is only relevant at values between 0 and 1, inclusive. This gives rise to the consideration of a transformation of our data to allow for expected values to be constrained

to be between 0 and 1. The logit transformation is such a transformation, which maps probabilities, p, measured between 0 and 1, onto a $-\infty$ to $+\infty$ scale. The logit transform is expressed then as a function of p:

$$g(p) = \ln [p/(1 - p)]$$

For a predictor x with probability of success $\pi(x) = E(Y|x)$,

$$g(x) = \ln [\pi(x)/(1 - \pi(x))] = \beta_0 + \beta_1 x$$

The logistic regression model then takes the form

$$\pi(x) = e^{\beta_0 + \beta_1 x} / \left(1 + e^{\beta_0 + \beta_1 x}\right)$$

When plotted against a range of x values, this model has an S-shaped curve, approaching values of 0 and 1 gradually as shown in Figure 24.1 for positive and negative values of β_1 (2).

The value of β_1 determines the rate of change in $\pi(x)$, with higher values indicating a faster rate of change. As in linear regression, a value of 0 for β_1 indicates that the response is independent of x.

Using the logistic regression model, the deviation between the expectation of a particular binary observation (the conditional mean, $E(Y|x)$) and the true value, 0 or 1, can be denoted by the random variable ε, as shown in the following equation: $y = E(Y|x) + \varepsilon$. Since y takes on only the values 0 and 1, ε can take on only two possible values: $1 - \pi(x)$ when $y = 1$ and $-\pi(x)$ when $y = 0$. Thus, the error term, ε, follows a binomial distribution with mean 0 and variance $\pi(x)[1 - \pi(x)]$ (14).

24.3   AN EXAMPLE UTILIZING SIMULATED CLINICAL TRIAL DATA

When modeling "real" data, an important first step is to carefully consider and specify the intended use of the model to be developed. With this in hand, an appropriate strategy and analysis plan can be crafted to ensure the appropriateness of the model (15). For the purposes of this chapter, a simulated data set is used to mimic a typical Phase 2 trial of a novel compound where an endpoint measurement is collected at each study visit. In this case, the endpoint of interest is the presence or absence of a particular adverse event. Various demographic variables are available for possible correlation with the endpoint in addition to a calculated measure of individual exposure to the drug.

In this example, a number of subjects receiving this particular compound in Phase 1 trials experienced rash. To further evaluate the potential relationship between exposure to the new drug and the probability of experiencing rash, a PK/PD logistic regression model is developed to explore and estimate this relationship. The simulated data are described in Table 24.1 and further explained in Section 24.3.2.2.

24.3.1   Exploratory Data Analysis

Prior to modeling, exploratory graphs and tables of the data to be analyzed should be generated to gain an understanding of the data that are to be modeled, to look for

FIGURE 24.1   Example logistic regression functions for positive and negative values of β_1.

trends in the data, to identify potential outliers or erroneous data values, to check for errors in coding that might have occurred during the data set creation process, and to verify model assumptions (16).

Contingency tables can easily be created to determine the frequencies of various combinations of variable levels. For instance, a contingency table of each discrete predictor versus the dependent variable can be created by summing up the numbers of 0 and 1 endpoints for each level of each discrete predictor. Such contingency tables are then examined to determine if there are any combinations of variable levels that result in a frequency of zero or a very small number of observations. If zero (or very small) frequencies are detected, one may consider either collapsing or combining two or more levels of the predictor variable or eliminating a particularly small category to avoid numerical problems during the modeling process. These tables are also useful as a univariate look for possible trends that may be observed in modeling (14). A sample contingency table for the rash data is provided in Table 24.2.

TABLE 24.1   Simulated Data Set Description

Position in Dataset Set	Variable Description	Coding	NONMEM Variable Name
1	Subject Identifier	ID Number	ID
2	Outcome = Rash	0 = Rash not observed between this and the prior visit 1 = Rash observed during the defined interval	DV
3	Drug Exposure	Calculated area under the concentration–time curve at steady state in ng·h/mL	AUC
4	NONMEM Missing Dependent Variable (MDV) Item	0 = DV not missing 1 = DV missing	MDV
5	Gender	0 = Male 1 = Female	SEXF
6	Study Visit	Integer value indicating study visit number	VIS
7	Patient Race	0 = Caucasian 1 = Black 2 = Hispanic 3 = Other	RACE
8	NONMEM Marginal Expectation Data Item	0 = No marginal expectation requested for this record 1 = Marginal expectation requested in PRED for this record	MRG_
9	NONMEM Raw Data Average Data Item	0 = No raw data average requested for this record 1 = Raw data average requested in DV for this record	RAW_

TABLE 24.2   Contingency Table for Endpoint and Race in the Rash Data Set

Frequency Row % Column %		Race Group				
		0 (Caucasian)	1 (Black)	2 (Hispanic)	3 (Other)	Total
Rash	0 (No)	1572	97	43	43	1755
		89.6	5.5	2.5	2.5	100.0
		83.4	85.1	78.2	91.5	83.5
	1 (Yes)	314	17	12	4	347
		90.5	4.9	3.5	1.2	100.0
		16.7	14.9	21.8	8.5	16.5
	Total	1886	114	55	47	2102
		89.7	5.4	2.6	2.2	100.0
		100.0	100.0	100.0	100.0	100.0

FIGURE 24.2 Empirical logit plotted versus the mean of the binned AUC values for each of 10 groups. A smooth line is drawn through the points to facilitate interpretation.

Since ordinary logistic regression models assume linearity between the logit and the predictor variables, this assumption should also be verified prior to modeling. Empirical logits can be calculated for each continuous predictor variable by ranking and "binning" the values of the continuous predictor and then calculating the value of the logit for that bin based on the number of successes observed in each bin and the total number of patients whose predictor falls into that bin. A plot of the empirical logit versus the mean value of the predictor in each bin should look reasonably linear if ordinary linear logistic regression is to be used. If this plot looks markedly nonlinear, either a transformation of the predictor variable should be considered or a nonlinear model should be utilized (2, 14). A plot of the empirical logit versus AUC for the example rash data set is shown in Figure 24.2. Given the reasonably linear relationship observed between the logit and AUC in this plot, as evidenced by the smooth line, this plot supports the assumption of linear logits for the AUC predictor.

In typical exposure–response (population PK/PD) evaluations, a variety of demographic, pathophysiologic, and laboratory variables are available for consideration as potential predictors of outcome. While serious consideration of the likelihood for a relationship as it relates to the compound's mechanism of action should be entertained before all possible variables are included for evaluation as predictors, the resulting list will usually still consist of a number of demographic variables that are likely to be correlated. As with any regression analysis, the correlations between potential predictors should be evaluated prior to inclusion in the model. Strong correlations between one or more potential predictors may result in multicollinearity in the model, making the assessment of the importance of predictors impossible to discern. Correlations between predictors of interest may warrant a decision to examine only one of the correlated variables as a possible predictor of outcome. Scatterplot matrices or pairwise scatterplots of the continuous predictor variables may be generated to examine these correlations.

24.3.2   Implementation and Coding of Logistic Regression Models in NONMEM

24.3.2.1   NM-TRAN Specifications

Once exploratory analyses are complete and the data set for analysis has been sufficiently examined for the appropriateness of the assumptions, a base structural model may be fit. Logistic regression models, as implemented in NONMEM, require the use of the Laplacian estimation method with the LIKELIHOOD option ($EST METH=COND LAPLAC LIKE). In general, when binary endpoint data (single or multiple responses collected from each individual) are modeled, only one level of random effects is estimated (usually OMEGA, added to the logit as a homoscedastic error, ETA(1)), representing the unexplained inter-subject variability in the predicted response. When only a single endpoint obser-vation is available from each patient, ETA(1) may still be added to the logit with the estimate of OMEGA fixed to 0 ($OMEGA 0 FIXED). The use of the LIKELIHOOD option with categorical data specifies to NONMEM to compute a different objective function than is computed for continuous data, one that is based on the conditional likelihood of η_j.

Since the model is user supplied, the $PRED block is used and only the NONMEM-required variables ID (subject identifier), DV (the dependent variable or dichotomous endpoint), MDV (the missing dependent variable data item), and the explanatory variables of interest (the predictors, or independent variables) are required. An example NM-TRAN control stream for the base model (described below) estimat-ing the probability of experiencing rash, prior to the inclusion of any potential covariate (predictor) effects, is shown in Appendix 24.1. In this case, the base model to be specified in NONMEM is described in the Eq. (24.1) and (24.2):

$$Logit_j = \theta_1 + \eta_j \tag{24.1}$$

$$P_j = e^{Logit_j} / \left(1 + e^{Logit_j}\right) \tag{24.2}$$

where θ_1 is the typical value of the logit; η_j is the discrepancy between the typical value of the logit and the true logit in the jth patient; η_j are independent and identically distributed random variables with mean 0 and variance ω^2; and P_j is the predicted probability of experiencing rash in the jth patient.

24.3.2.2   Data Set Specifications

A listing of the data for the first and last two patients in the data set is provided in Table 24.3. The DV data item is used to denote whether (a value of 1) or not (a value of 0) rash was observed in the given patient. Most other variables and values are self-explanatory with the exception of the MRG_ and RAW_, the marginal expec-tation and raw data average data items. These data items are indicator variables for NONMEM to signal the calculation of various statistics for the given values of other variables on the record. The MRG_ data item with a value of 1 indicates to set PRED (the NONMEM-generated typical value prediction) to the expected value of Y. MRG_ can be used to generate both simulation expectations (in simulation problems) and posterior expectations (in estimation problems or following estima-tion using $MSFI). The RAW_ data item with a value of 1 indicates to set DV to the

**TABLE 24.3   Records for the First and Last Two Patients of
NONMEM-Formatted Data Set for Example**

ID	DV	AUC	MDV	SEXF	VIS	RACE	MRG_	RAW_
1	0	21.99	0	0	1	1	0	0
1	0	41.54	0	0	2	1	0	0
1	0	81.00	0	0	3	1	0	0
1	0	158.91	0	0	4	1	0	0
1	0	316.18	0	0	5	1	0	0
1	0	472.72	0	0	6	1	0	0
1	1	629.99	0	0	7	1	0	0
1	0	629.99	0	0	8	1	0	0
1	1	940.19	0	0	9	1	0	0
1	0	940.19	0	0	10	1	0	0
1	0	940.19	0	0	11	1	0	0
1	0	940.19	0	0	12	1	0	0
2	0	25.67	0	0	1	1	0	0
—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—
998	1	18.28	1	0	1	0	1	1
—	—	—	—	—	—	—	—	—
998	1	85.97	1	0	1	0	1	1
999	1	19.64	1	1	1	0	1	1
—	—	—	—	—	—	—	—	—
999	1	925.06	1	1	1	0	1	1

raw data average for the given value(s) of the other data items. MRG_ and RAW_ can be used together (as shown in this example) to generate both expectations and averages for the values of the variables on these records (in our data set, ID = 998 and 999) (1).

24.3.2.3   Interpretation of the Output

The report file from the base model run provided a minimum value of the objective function of 1616.814 and an estimate (SE) for θ_1 of −2.43 (0.173) and ω^2_1 of 2.88 (0.538). The marginal posterior expectation is 0.155 over all values of AUC and other covariates (none of which are included in the model yet) and the raw data average is 0.162 (approximately equivalent to the frequency of occurrences of rash in the data set divided by the total number of patients).

To get a sense of the goodness of fit for a particular model, a plot of the raw data average (values in DV for RAW_ = 1) and the posterior expectation (values in PRED for MRG_ = 1) versus each other and/or explanatory variables can be generated (see Section 24.4 for an example of this plot). Another assessment of model fit can be obtained by plotting the individual η values versus potential predictors of interest. Trends observed in a plot of η versus a covariate may indicate a possible relationship between the outcome and that predictor which has not yet been accounted for in the model.

24.4 MODEL BUILDING

With the base model fit in hand, any of a number of different strategies may be employed to evaluate the influence of the exposure variables and covariates on the response. As with other population PK (and PK/PD or PD) analyses, many different techniques and processes have been advocated for efficiently and effectively screening and selecting covariates for inclusion in a model (17–19). For the purposes of this chapter, the model including the effect of exposure (AUC) on the response is illustrated, as is the final model, including other covariate effects (presumably derived following the application of some technique to screen all potential covariates).

Given the apparent linearity observed in the plot of the empirical logit versus AUC (see Figure 24.2), a linear model including (a centered effect of) AUC is evaluated. This model may be specified as

$$Logit_j = \theta_1 + \theta_2(AUC_j - 118) + \eta_j \tag{24.3}$$

$$P_j = e^{Logit_j}/(1 + e^{Logit_j}) \tag{24.4}$$

where θ_1 is the typical value of the logit at the median AUC value of $118\,ng\cdot h/mL$; θ_2 is the typical value of the slope relating AUC to the logit; AUC_j is the calculated AUC in the jth patient (centered about the population median value of $118\,ng\cdot h/mL$); η_j is the discrepancy between the typical value of the logit and the true logit in the jth patient; η_j are independent and identically distributed random variables with mean 0 and variance ω^2; and P_j is the predicted probability of experiencing rash in the jth patient.

To compare this model to the simpler base model, a likelihood ratio test may be utilized as is commonly applied in population model building. This test considers the log likelihood values (in NONMEM, the minimum values of the objective function) from two hierarchical models and compares the difference in these values to a χ^2 statistic with the number of degrees of freedom equal to the difference in the number of parameters estimated in the two models. When the model including the effect of AUC was estimated in the example rash data set, the minimum value of the objective function was 1605.344. Thus, the difference in the log-likelihood values for the two models is 11.470 with 1 degree of freedom, relating to a p-value of 0.0007, and the conclusion that AUC is a statistically significant predictor of response at $\alpha = 0.05$. The estimates (SE) for θ_1 and θ_2 were –2.54 (0.176) and 0.000969 (0.000287), respectively, and ω^2_1 was estimated at 2.80 (0.525).

A commonly reported statistic for logistic regression models is the odds ratio. An odds ratio, or the ratio of the odds for $x = 1$ to the odds for $x = 0$, may be calculated (14) as

$$\Psi = \frac{[\pi(1)/(1-\pi(1))]}{[\pi(0)/(1-\pi(0))]} \tag{24.5}$$

For logistic regression models with a single predictor variable, x, this equation reduces to e^{β_1}. For a dichotomous predictor such as gender, this ratio approximates how much more likely (or unlikely) it is for the endpoint of interest to be observed for females as compared to males. For a continuous predictor such as exposure,

this ratio approximates how much more likely (or unlikely) it is for the endpoint of interest to be observed for every one-unit increase in exposure. Since a 1 ng·h/mL increase in exposure may not be particularly meaningful depending on the scale and range of the predictor, odds ratios for continuous predictors can also be calculated for larger increments in the predictor (c) to facilitate interpretation by multiplying β_1 by c before exponentiating (14).

This ratio also provides another means of interpreting the coefficient for the predictor x. We can say that the odds of experiencing the endpoint of interest increase by e^{β_1} times for every one-unit increase in x. A 95% confidence interval around this odds ratio can be calculated by first determining the interval around β_1 and then exponentiating the endpoints (for an arbitrary unit increment of c):

$$e^{[c\cdot\beta_1 - z_{1-\alpha/2}\cdot c\cdot SE(\beta_1)]}, \quad e^{[c\cdot\beta_1 + z_{1-\alpha/2}\cdot c\cdot SE(\beta_1)]} \tag{24.6}$$

Therefore, given the estimate of θ_2 (or β_1) in the exposure–response model, AUC is associated with an odds ratio of $e^{\beta_1} = e^{0.000969} = 1.001$. The 95% confidence interval associated with this odds ratio is

$$(e^{\beta_1 - 1.96\cdot SE(\beta_1)}, e^{\beta_1 + 1.96\cdot SE(\beta_1)}) = (e^{0.000969 - 1.96\cdot 0.000287}, e^{0.000969 + 1.96\cdot 0.000287}) = (1.0004, 1.0015)$$

This odds ratio for exposure means that the odds of experiencing rash increase by 0.1% with every 1 ng·h/mL increase in AUC over the median value of 118 ng·h/mL. Alternatively, if a 100 ng·h/mL increase in AUC was thought to be a meaningful difference, the resulting odds ratio would indicate that the odds of experiencing rash increase by $e^{100\cdot 0.000969} = 1.10$ times for every 100 ng·h/mL increase in AUC over the median value of 118 ng·h/mL. For a patient with the lowest calculated AUC (8.38 ng·h/mL), this translates into a predicted probability (marginal posterior expectation) of rash of $e^{-2.54+0.000969\cdot(8.38-118)}/(1 + e^{-2.54+0.000969\cdot(8.38-118)}) = 0.066$ and for a patient with the highest calculated AUC (1603.88 ng·h/mL), a probability of $e^{-2.54+0.000969\cdot(1603.88-118)}/(1 + e^{-2.54+0.000969\cdot(1603.88-118)}) = 0.250$.

Another statistic that can be calculated for the logistic regression model is the median effective level, or EL_{50}; in our example, this is interpreted as the AUC at which there is a 50% probability of either experiencing or not experiencing the endpoint (rash). The EL_{50} can be estimated by $-\beta_0/\beta_1$, the x value associated with the steepest slope of the S-shaped logistic regression curve. For this model, the EL_{50} is the AUC at which the predicted probability is 0.5, which is $-\beta_0/\beta_1 = 2.54/0.000969 = 2621.26$ ng·h/mL. While this EL_{50} value is well outside the range of the observed AUCs in this data set (maximum AUC = 1603.9 ng·h/mL), this is not unexpected due to the low incidence of rash in these simulated trial data. For a more prevalent endpoint, this statistic may be useful in interpreting the parameter estimates.

Following covariate analysis for this simulated data set, gender was the only covariate determined to be a statistically significant predictor of response in addition to exposure at a significance level of $\alpha = 0.05$ for entrance into the model. The only other covariate tested in this example was race, and the model including exposure and race did not result in a statistically significant change in the minimum value of the objective function. The final model, including the effects of exposure and gender, is described in the following equations:

$$Logit_j = \theta_1 + \theta_2 \cdot (AUC_j - 118) + \theta_3 \cdot SEXF_j + \eta_j \qquad (24.7)$$

$$P_j = e^{Logit_j} / \left(1 + e^{Logit_j}\right) \qquad (24.8)$$

where θ_1 is the typical value of the logit for a male patient at the median AUC value of 118 ng·h/mL; θ_2 is the typical value of the slope relating AUC to the logit; AUC_j is the calculated AUC in the jth patient (centered about the population median value of 118 ng·h/mL); θ_3 is the typical value of the increment (or decrement) in the logit relating to female gender; $SEXF_j$ is an indicator variable for gender in the jth patient (with 0 = male and 1 = female); η_j is the discrepancy between the typical value of the logit and the true logit in the jth patient; η_j are independent and identically distributed random variables with mean 0 and variance ω^2; and P_j is the predicted probability of experiencing rash in the jth patient.

When the model including the effect of gender was estimated in the example rash data set, the minimum value of the objective function was 1597.624. Thus, the difference in the log-likelihood values between this model and the exposure–response model not including the effect of gender is 7.72 with 1 degree of freedom, relating to a p-value of 0.005, and the conclusion that gender is a statistically significant predictor of response at $\alpha = 0.05$. The estimates (SE) for θ_1, θ_2, and θ_3 were −3.00 (0.256), 0.000961 (0.000286), and 0.824 (0.299), respectively, and $\omega^2{}_1$ was estimated at 2.69 (0.510). Given the estimate of θ_3 (or β_2), female gender is associated with an odds ratio (95% confidence interval) of 2.28 (1.27, 4.10). This odds ratio for gender means that, after accounting for exposure, the odds of experiencing rash are increased by approximately two times for females as compared to males. This translates into predicted probabilities (marginal posterior expectations) of rash for male patients ranging from 0.043 to 0.119 and for female patients from 0.093 to 0.321 for the lowest calculated AUC to the highest calculated AUC values for each gender (8.38 and 1156.61 and 10.16 and 1603.88 ng·h/mL, respectively). Figure 24.3 shows a plot of the raw data average and the posterior expectation versus the mean of each AUC bin separately for male and female patients.

When exploring the influence of race on the probability of experiencing rash, the model failed to converge without error. Since the exploratory analysis revealed that some of the race groups contained very small numbers of patients experiencing rash, after evaluating the model including a separate indicator variable for each race group, a model with all race groups other than Caucasian combined into one group was also estimated. Although the run including only one race effect resulted in a successful minimization, it still failed to reach statistical significance for inclusion of the effect in the model. A NM-TRAN control stream for the final model estimating the probability of experiencing rash, as a function of both exposure (AUC) and gender, is shown in Appendix 24.2. The associated report file for this final model is provided in Appendix 24.3.

24.5   MODEL EXTENSIONS

If the binary endpoints available from each patient were collected at various times or visits and there was interest in estimating the effect of time or visit on response, a mixed effects longitudinal logistic regression model could be utilized. Longitudinal logistic regression models can be thought of as similar to the logistic regression

FIGURE 24.3 Data-based probabilities and posterior expectations versus the mean of the binned AUC values for male and female patients. In each plot, the data-based probabilities, or raw data averages (asterisks and open circles), are joined by a dashed smooth line through the points. The posterior expectations (open triangles and filled circles) are joined by a solid line illustrating the model-based predictions for the mean value of each AUC bin.

models we have studied with an additional covariate estimating the fixed effect relating to time. If the values of other patient covariates also change over time, these changing values can be incorporated into the NONMEM data set much as they would in a typical population PK analysis with time-varying covariates.

Covariate interactions, if biologically plausible, can also be estimated and tested for significance as they would with other linear or mixed effects models. To determine whether an interaction between covariates might be present, the empirical logit plots described in Section 24.3.1 can be examined. To assess the possibility for continuous–dichotomous predictor interactions, the empirical logit can be plotted versus each continuous predictor with separate symbols and lines for each level

of the dichotomous predictors. For interactions between continuous predictors, binning of the values for one predictor can be used to create levels by which the logit versus the second predictor plot can be stratified. If these plots reveal intersecting trend lines for the levels of one covariate across another, evaluation of the interaction may be warranted.

A markedly nonlinear trend in the exploratory plot of the empirical logit versus exposure or a covariate of interest would indicate that the assumptions of the linear logistic model are not met. In this case, a more complex nonlinear model can be evaluated using a slight modification to the models already presented. Let's suppose that the plot of the empirical logit versus AUC had a distinctly concave trend, curving upward similarly at each end of the observed AUC distribution. With appropriate precautions and care in interpretation, especially outside the range of the data, a quadratic model may be implemented:

$$\pi(x) = e^{\beta_0 + \beta_1 x + \beta_2 x^2} \big/ \left(1 + e^{\beta_0 + \beta_1 x + \beta_2 x^2}\right) \tag{24.9}$$

On occasion, even with a large sample size, a situatin may arise where there are relatively few failures (or successes) or where all or none of the successes fall in one category of an important covariate (perfect discrimination) (2). In order to avoid erroneous conclusions in these situations, it is advisable to have a solid understanding of the data to be modeled through the creation of exploratory graphs and tables, as different software packages provide differing results when this condition arises. Exact methods can be used to compute parameter estimates and confidence intervals for very small sample size problems (20). However, specific point and interval estimates may not always be needed and a graphical and tabular illustration of the data may suffice in describing the apparent relationship between the endpoint and the covariates of interest.

24.6   SUMMARY

This chapter describes the theoretical basis for logistic regression analysis of binary endpoint data in PK/PD exposure–response-type assessments. Some important considerations in the implementation and coding of this technique using NONMEM are illustrated. A simulated data set is used to describe the recommended premodeling exploratory looks at the data, as well as the approach to model development, assessing goodness of fit, and covariate analysis considerations for dichotomous endpoint data. Several special features of NONMEM designed to facilitate the analysis of binary data, including data items to request the calculation of raw data averages and marginal posterior expectations and estimation options for the calculation of conditional expectations, are described. The use of NONMEM for the implementation of ordinary fixed and mixed effects logistic regression models is shown to be advantageous in adding extensions to the model to test for and estimate nonlinear functions, interactions between covariate effects, or other more complicated functions within the same general framework. In conclusion, binary outcome data routinely collected during clinical trials evaluating PK/PD relationships can be successfully analyzed using logistic regression techniques. Subsequent chapters consider more complex endpoints.

REFERENCES

1. Intermediate workshop in population pharmacokinetic data analysis using the NONMEM system. Regents of the University of California, 1992.

2. A. Agresti, *An Introduction to Categorical Data Analysis*. Wiley, Hoboken, NJ, 1996.

3. T. N. Johnson, A. Rostami-Hodjegan, J. M. Goddard, M. S. Tanner, and G. T. Sucker, Contribution of midazolam and its 1-hydroxy metabolite to preoperative sedation in children: a pharmacokinetic–pharmacodynamic analysis. *Br J Anaesth* **89**(3):428–437 (2002).

4. J. A. Johnston, J. Fiedler-Kelly, E. D. Glover, D. P. Sachs, T. H. Grasela, and J. DeVeaugh-Geiss, Relationship between drug exposure and the efficacy and safety of bupropion sustained release for smoking cessation. *Nicotine Tob Res* **3**:131–140 (2001).

5. D. Mould, M. Chapelshy, J. Aluri, J. Swagzdis, R. Samuels, and J. Granett, A population pharmacokinetic–pharmacodynamic and logistic regression analysis of lotrafiban in patients. *Clin Pharmacol Ther* **69**(4):210–222 (2001).

6. R. Xie, R. H. Mathijssen, A. Sparreboom, J. Verweij, and M. O. Karlsson, Clinical pharmacokinetics of irinotecan and its metabolites in relation with diarrhea. *Clin Pharmacol Ther* **72**(3):265–275 (2002).

7. *NONMEM Users Guides*, S. L. Beal and L. B . Sheiner (Eds.). NONMEM Project Group, University of California at San Francisco, San Francisco, 1992.

8. SAS Institute Inc., *SAS/STAT User's Guide*, *Version 8*, Vols. 1–3. SAS Institute Inc., Cary, NC, 1999, p. 3809.

9. A. Hsu, J. Isaacson, S. Brun, B. Bernstein, W. Lam, R. Bertz, C. Foit, K. Rynkiewicz, B. Richards, M. King, R. Rode, D. J. Kempf, G. R. Granneman, and E. Sun, Pharmacokinetic–pharmacodynamic analysis of lopinavir–ritonavir in combination with efavirenz and two nucleoside reverse transcriptase inhibitors in extensively pretreated human immunodeficiency virus-infected patients. *Antimicrob Agents Chemother* **47**(1):350–359 (2003).

10. S. Weller, K. M. Radomski, U. Lou, and D. S. Stein, Population pharmacokinetics and pharmacodynamic modeling of abacavir (1592U89) from a dose-ranging, double-blind, randomized monotherapy trial with human immunodeficiency virus-infected subjects. *Antimicrob Agents Chemother* **44**(8):2052–2060 (2000).

11. R. Gieschke, H. U. Burger, B. Reigner, K. S. Blesch, and J. L. Steimer, Population pharmacokinetics and concentration–effect relationships of capecitabine metabolites in colorectal cancer patients. *Br J Clin Pharmacol* **55**(3):252–263 (2003).

12. I. Yano, S. L. Beal, and L. B. Sheiner, The need for mixed-effects modeling with population dichotomous data. *J Pharamacokinet Pharmacodyn* **28**(4):389–412 (2001).

13. D. Collett, *Modelling Binary Data*, 2nd ed. Chapman & Hall/CRC, New York, 2003.

14. D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, Hoboken, NJ, 1989.

15. E. I. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M. J. Liu, and E. V. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).

16. *FDA Guidance for Industry, Population Pharmacokinetics.* http://www.fda.gov/cder/guidance/1852fnl.pdf.

17. J. W. Mandema, D. Verotta, and L. B. Sheiner, Building population pharmacokinetic–pharmacodynamic models. I. Models for covariate effects. *J Pharmacokinet Biopharm* **20**(5):511–528 (1992).

18. E. N. Jonsson and M. O. Karlsson, Automated covariate model building within NONMEM. *Pharm Res* **15**(9):1463–1468 (1998).

19. K. G. Kowalski and M. M. Hutmacher, Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *J Pharamacokinet Pharmacodyn* **28**(3):253–275 (2001).

20. C. R. Mehta and N. R. Patel, Exact logistic regression: theory and examples. *Stat Med* **14**:2143–2160 (1995).

APPENDIX 24.1   NONMEM CODE FOR BASE MODEL EXAMPLE

```
$PROB Rash example - base model run
$DATA CH24EXAMPLE.csv IGNORE=C
; The data can be found in the Excel file, CH24EXAMPLE.csv
$INPUT ID DV AUC MDV SEXF VIS RACE MRG_ RAW_
$PRED

LOGIT = THETA(1) + ETA(1)
A = EXP(LOGIT)
P = A / (1 + A)

IF (DV .EQ. 1) Y=P
IF (DV .EQ. 0) Y=1-P

$THETA (-2)
$OMEGA 5
$ESTIMATION METHOD=COND LAPLACE LIKELIHOOD MAXEVAL=5000 PRINT=2
  MSFO=../base.msf
$COV MATRIX = R
$TABLE ID SEXF AUC RACE VIS ETA1 NOPRINT NOHEADER FILE=../base.
tbl
$SCAT ETA1 VS (AUC RACE SEXF) FROM 1 TO 900
$SCAT ETA1 VS (AUC RACE SEXF) FROM 901 TO 1800
$SCAT ETA1 VS (AUC RACE SEXF) FROM 1801
```

APPENDIX 24.2   NONMEM CODE FOR FINAL MODEL EXAMPLE

```
$PROB Rash example - E-R model run with gender effect
$DATA CH24EXAMPLE.csv IGNORE=C
$INPUT ID DV AUC MDV SEXF VIS RACE MRG_ RAW_
$PRED

LOGIT = THETA(1) + THETA(2)*(AUC-118) + THETA(3)*SEXF + ETA(1)
A = EXP(LOGIT)
P = A / (1 + A)
IF (DV .EQ. 1) Y=P
IF (DV .EQ. 0) Y=1-P
$THETA (-2) (0.0007) (0.4)
$OMEGA 5
$ESTIMATION METHOD=COND LAPLACE LIKELIHOOD MAXEVAL=5000 PRINT=2
```

```
  MSFO=../ersexf.msf
$COV MATRIX=R
$TABLE ID SEXF AUC RACE VIS ETA1 NOPRINT NOHEADER FILE=../ersexf.
tbl
$SCAT RES VS (AUC SEXF) FROM 1 TO 900
$SCAT RES VS (AUC SEXF) FROM 901 TO 1800
$SCAT RES VS (AUC SEXF) FROM 1801
$SCAT PRED VS (AUC SEXF) FROM 1 TO 900
$SCAT PRED VS (AUC SEXF) FROM 901 TO 1800
$SCAT PRED VS (AUC SEXF) FROM 1801
$SCAT ETA1 VS (AUC RACE SEXF) FROM 1 TO 900
$SCAT ETA1 VS (AUC RACE SEXF) FROM 901 TO 1800
$SCAT ETA1 VS (AUC RACE SEXF) FROM 1801
```

APPENDIX 24.3   NONMEM REPORT FILE FOR FINAL MODEL EXAMPLE

```
1NONLINEAR MIXED EFFECTS MODEL PROGRAM (NONMEM) DOUBLE PRECISION
NONMEM VERSION V LEVEL 1.1
  DEVELOPED AND PROGRAMMED BY STUART BEAL AND LEWIS SHEINER

  PROBLEM NO.: 1
  Rash example - E-R model run with gender effect
0DATA CHECKOUT RUN:                 NO
  DATA SET LOCATED ON UNIT NO.:      2
  THIS UNIT TO BE REWOUND:          NO
  NO. OF DATA RECS IN DATA SET:   2122
  NO. OF DATA ITEMS IN DATA SET:     9
  ID DATA ITEM IS DATA ITEM NO.:     1
  DEP VARIABLE IS DATA ITEM NO.:     2
  MDV DATA ITEM IS DATA ITEM NO.:    4
  MRG DATA ITEM IS DATA ITEM NO.:    8
  RAW DATA ITEM IS DATA ITEM NO.:    9
0LABELS FOR DATA ITEMS:
  ID DV AUC MDV SEXF VIS RACE MRG_ RAW_
0FORMAT FOR DATA:
  (9E8.0)

  TOT. NO. OF OBS RECS:          2102
  TOT. NO. OF INDIVIDUALS:        202
0LENGTH OF THETA: 3
0OMEGA HAS SIMPLE DIAGONAL FORM WITH DIMENSION: 1
0INITIAL ESTIMATE OF THETA:
 -0.2000E+01 0.7000E-03 0.4000E+00
0INITIAL ESTIMATE OF OMEGA:
 0.5000E+01
0ESTIMATION STEP OMITTED:          NO
  CONDITIONAL ESTIMATES USED:      YES
  CENTERED ETA:                    NO
  EPS-ETA INTERACTION:             NO
```

```
 LAPLACIAN OBJ. FUNC.:              YES
 NUMERICAL 2ND DERIVATIVES:          NO
 PRED F SET TO A LIKELIHOOD:        YES
 NO. OF FUNCT. EVALS. ALLOWED:     5000
 NO. OF SIG. FIGURES REQUIRED:        3
 INTERMEDIATE PRINTOUT:             YES
 ESTIMATE OUTPUT TO MSF:            YES
0COVARIANCE STEP OMITTED:            NO
 R MATRIX SUBSTITUTED:              YES
 S MATRIX SUBSTITUTED;               NO
 EIGENVLS. PRINTED:                  NO
 COMPRESSED FORMAT:                  NO
0TABLES STEP OMITTED:                NO
 NO. OF TABLES:                       1
0- TABLE 1 -
 PRINTED:                            NO
 HEADERS:                            NO
 FILE TO BE FORWARDED:               NO
0USER-CHOSEN ITEMS
  IN THE ORDER THEY WILL APPEAR IN THE TABLE:
  ID SEXF AUC RACE VIS ETA1
0SCATTERPLOT STEP OMITTED:           NO
 FAMILIES OF SCATTERPLOTS:           18
0- SCATTERPLOT 1 -
 UNIT SLOPE LINE:                    NO
 BEGINNING DATA REC.:                 1
 ENDING DATA REC.:                  900
0ITEMS TO BE SCATTERED:             AUC    RES
0- SCATTERPLOT 2 -
 UNIT SLOPE LINE:                    NO
 BEGINNING DATA REC.:                 1
 ENDING DATA REC.:                  900
0ITEMS TO BE SCATTERED:            SEXF    RES
0- SCATTERPLOT 3 -
 UNIT SLOPE LINE:                    NO
 BEGINNING DATA REC.:               901
 ENDING DATA REC.:                 1800
0ITEMS TO BE SCATTERED:             AUC    RES
0- SCATTERPLOT 4 -
 UNIT SLOPE LINE:                    NO
 BEGINNING DATA REC.:               901
 ENDING DATA REC.:                 1800
0ITEMS TO BE SCATTERED:            SEXF    RES
0- SCATTERPLOT 5 -
 UNIT SLOPE LINE:                    NO
 BEGINNING DATA REC.:              1801
0ITEMS TO BE SCATTERED:          AUC RES
0- SCATTERPLOT 6 -
 UNIT SLOPE LINE:                    NO
```

```
  BEGINNING DATA REC.:            1801
0ITEMS TO BE SCATTERED:          SEXF    RES
0- SCATTERPLOT 7 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             1
  ENDING DATA REC.:               900
0ITEMS TO BE SCATTERED:          AUC    PRED
0- SCATTERPLOT 8 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             1
  ENDING DATA REC.:               900
0ITEMS TO BE SCATTERED:          SEXF    PRED
0- SCATTERPLOT 9 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:            901
  ENDING DATA REC.:               1800
0ITEMS TO BE SCATTERED:          AUC    PRED
0- SCATTERPLOT 10 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:            901
  ENDING DATA REC.:               1800
0ITEMS TO BE SCATTERED:          SEXF    PRED
0- SCATTERPLOT 11 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:            1801
0ITEMS TO BE SCATTERED:          AUC    PRED
0- SCATTERPLOT 12 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:            1801
0ITEMS TO BE SCATTERED:          SEXF    PRED
0- SCATTERPLOT 13 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             1
  ENDING DATA REC.:               900
0ITEMS TO BE SCATTERED:          AUC    ETA1
0- SCATTERPLOT 14 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             1
  ENDING DATA REC.:               900
0ITEMS TO BE SCATTERED:          RACE    ETA1
0- SCATTERPLOT 15 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             1
  ENDING DATA REC.:               900
0ITEMS TO BE SCATTERED:          SEXF    ETA1
0- SCATTERPLOT 16 –
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:            901
  ENDING DATA REC.:               1800
```

```
0ITEMS TO BE SCATTERED:           AUC    ETA1
0- SCATTERPLOT 17 -
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             901
  ENDING DATA REC.:               1800
0ITEMS TO BE SCATTERED:           RACE   ETA1
0- SCATTERPLOT 18 -
  UNIT SLOPE LINE:                 NO
  BEGINNING DATA REC.:             901
  ENDING DATA REC.:               1800
0ITEMS TO BE SCATTERED:           SEXF   ETA1
1
  MONITORING OF SEARCH:


0ITERATION NO.: 0 OBJECTIVE VALUE: 0.1642E+04 NO. OF FUNC. EVALS.: 5
  CUMULATIVE NO. OF FUNC. EVALS.:  5
  PARAMETER: -0.1000E+00   0.1000E+00   0.1000E+00 0.1000E+00
  GRADIENT:   0.1184E+04   0.9634E+01   0.1047E+03 0.8311E+03
0ITERATION NO.: 2 OBJECTIVE VALUE: 0.1601E+04 NO. OF FUNC. EVALS.:10
  CUMULATIVE NO. OF FUNC. EVALS.: 23
  PARAMETER: -0.1302E+00   0.1017E+00   0.9906E-01 0.7231E-01
  GRADIENT:   0.1842E+03  -0.3523E+02  -0.1847E+02 0.7041E+02
0ITERATION NO.: 4 OBJECTIVE VALUE: 0.1598E+04 NO. OF FUNC. EVALS.: 7
  CUMULATIVE NO. OF FUNC. EVALS.: 37
  PARAMETER: -0.1448E+00   0.1362E+00   0.2061E+00 0.7213E-01
  GRADIENT:   0.1616E+03   0.6380E+01   0.1962E+02 0.3591E+02
0ITERATION NO.: 6 OBJECTIVE VALUE: 0.1598E+04 NO. OF FUNC. EVALS.: 6
  CUMULATIVE NO. OF FUNC. EVALS.: 49
  PARAMETER: -0.1498E+00   0.1371E+00   0.2060E+00 0.7351E-01
  GRADIENT:   0.1281E+01  -0.2471E+00  -0.1134E+00 0.4347E+01
0ITERATION NO.: 8 OBJECTIVE VALUE: 0.1598E+04 NO. OF FUNC. EVALS.: 7
  CUMULATIVE NO. OF FUNC. EVALS.: 62
  PARAMETER: -0.1499E+00   0.1373E+00   0.2066E+00 0.7343E-01
  GRADIENT:  -0.2672E+01  -0.7355E-01  -0.7572E-01 0.1853E-01
0ITERATION NO.: 10 OBJECTIVE VALUE: 0.1598E+04 NO. OF FUNC. EVALS.: 0
  CUMULATIVE NO. OF FUNC. EVALS.: 71
  PARAMETER: -0.1498E+00   0.1373E+00   0.2060E+00 0.7340E-01
  GRADIENT:  -0.6623E+00  -0.1218E-01  -0.1736E-01 0.1033E+00
0MINIMIZATION SUCCESSFUL
  NO. OF FUNCTION EVALUATIONS USED: 71
  NO. OF SIG. DIGITS IN FINAL EST.: 3.0
  ETABAR IS THE ARITHMETIC MEAN OF THE ETA-ESTIMATES,
  AND THE P-VALUE IS GIVEN FOR THE NULL HYPOTHESIS THAT THE TRUE
  MEAN IS 0.

ETABAR: 0.20E+00
P VAL.: 0.20E-01
1
```

```
*********************************************************************
   *********************
                *********************
   *********************        MINIMUM  VALUE  OF  OBJECTIVE  FUNCTION
                *********************
   *********************
                *********************
*********************************************************************
   ******************************************************    1597.624
***********************************************
1
*********************************************************************
   *********************
                *********************
   *********************                         FINAL  PARAMETER  ESTIMATE
                *********************
   *********************
                *********************
*********************************************************************
THETA - VECTOR OF FIXED EFFECTS PARAMETERS *********
  TH 1      TH 2      TH 3
  -3.00E+00 9.61E-04 8.24E-01
OMEGA - COV MATRIX FOR RANDOM EFFECTS - ETAS ********
  ETA1
ETA1
+ 2.69E+00
1
*********************************************************************
   *********************
                *********************
   *******************                           STANDARD ERROR OF ESTIMATE
                *********************
   *********************
                *********************
*********************************************************************
THETA - VECTOR OF FIXED EFFECTS PARAMETERS *********
  TH 1      TH 2      TH 3
  2.56E-01 2.86E-04 2.99E-01
OMEGA - COV MATRIX FOR RANDOM EFFECTS - ETAS ********
  ETA1
ETA1
+ 5.10E-01
1
*********************************************************************
   *********************
                *********************
*******************                           COVARIANCE MATRIX OF ESTIMATE
                *********************
```

```
   ********************
               ********************
*****************************************************************
  TH 1       TH 2      TH 3      OM11
TH 1
+ 6.56E-02
TH 2
+ -1.13E-05  8.19E-08
TH 3
+ -5.64E-02 -2.89E-07 8.95E-02
OM11
+ -5.66E-02  2.89E-06 2.12E-02 2.60E-01
1
*****************************************************************
  ********************
               ********************
  ********************                CORRELATION  MATRIX  OF  ESTIMATE
               ********************
  ********************
               ********************
*****************************************************************
  TH 1       TH 2      TH 3      OM11
TH 1
+  1.00E+00
TH 2
+ -1.55E-01  1.00E+00
TH 3
+ -7.36E-01 -3.37E-03 1.00E+00
OM11
+ -4.33E-01  1.98E-02 1.39E-01 1.00E+00
1
*****************************************************************
  ********************
               ********************
  ********************          INVERSE COVARIANCE MATRIX OF ESTIMATE
               ********************
  ********************
               ********************
*****************************************************************
  TH 1       TH 2      TH 3      OM11
TH 1
+ 4.71E+01
TH 2
+ 6.33E+03 1.31E+07
TH 3
+ 2.78E+01 3.81E+03 2.78E+01
OM11
+ 7.90E+00 9.22E+02 3.74E+00 5.25E+00
```

Population Pharmacokinetic/ Pharmacodynamic Modeling of Ordered Categorical Longitudinal Data

ENE I. ETTE, AMIT ROY, and PARTHA NANDY

25.1 INTRODUCTION

A variable that has a measurement scale consisting of a set of categories is termed a categorical variable. Endpoints commonly used in clinical studies or clinical practice are usually ranked ordered and are therefore ordered categorical in nature. Terms such as mild, moderate, or severe are used to describe adverse events, and different ranking scales are usually used for efficacy measures in clinical trials. The latter range from classifying subjects in an efficacy trial as responders or nonresponders— a binary outcome—to outcomes measured on an ordered scale. An example of a binary outcome would be treatment failure or success. Outcomes measured on an ordered categorical scale with several levels could include adverse events, sedation (1, 2), and pain scores (3–8), among others.

Two types of scales are primarily used for measuring categorical variables. These are the nominal and ordinal scales (9). Variables such as race and gender that have category without a natural ordering are nominal variables, and for these variables the order of listing of the categories is irrelevant. Analysis of such data does not depend on their ordering. On the other hand, many categorical variables have ordered categories and such variables are called ordinal variables. Some examples of these are pain scores, adverse events (mild, moderate, severe, and life threatening) discussed in the previous paragraph. Most outcome variables in clinical trials are measured as ordinal variables, and they can be analyzed with marginal and/or conditional models depending on the objective of the analysis. Marginal models are empirical models that characterize population-averaged effects, and generalized estimation equations are an example of such models. Conditional models, on the other hand, characterize subject-specific effects as well as population-averaged effects. A good example of these types of models is a mixed effects model. In longitudinal (repeated measure) clinical trials, categorical pharmacodynamic

(PD) (i.e., efficacy and safety) data are collected over the entire duration of the trial from each subject. Analyzing data from such trials to extract all the knowledge (subject-specific and population-averaged) contained in the data would require the use of a conditional model. A generalization of the logistic model to a multiple-category response model is the most popular model for ordered categorical data (10). The logit function is used as a link function to link the observations to the cumulative probabilities of the parameters of this model (11). It is possible to use the model with another link function other than the logit (e.g., the probit). Provision can be made for separate effects and use of the partial proportional odds model (12). Sheiner (3) used a conditional model to analyze ordinal data (pain scores). Thus, he pioneered the use of the proportional odds model with mixed effects in the analysis of analgesic trial data with nonrandom censoring (see below).

25.2 SURVIVAL DATA

An outcome of interest in many clinical trials is the time to an event. The time to the occurrence of an event is termed survival time. Examples of such events include death, the time it takes for a patient to respond to a therapy, the time to relapse after having responded to therapy, the time to tumor progression, and the time to rescue therapy in an analgesic trial. One may be interested in characterizing the distribution of time to an event for a given population, in comparing the time to an event among different treatment groups, or in modeling the relationship between the time to an event and subject-specific covariates. Thus, survival data include survival time, survival, time to analgesic remedication, response to a given treatment, subject characteristics related to response, and time to the development of a disease, among others. The implementation of survival analysis in the nonlinear mixed effects modeling setting has been shown to be useful in constructing subject-specific dose–response curves from analgesic trials, comparing dosage forms of an analgesic, and designing dosage regimens (3–8).

   In the sections that follow, terminologies and functions used to characterize survival data are first explained, followed by the application of nonlinear mixed effects modeling to the analysis of nonrandomly censored ordered categorical longitudinal data with application to analgesic trials.

25.2.1 Censored Data

The distinguishing characteristic of survival data is that the exact time to event is usually not known for all the subjects in the study. The most common reason for this is that the event may not have occurred in all subjects by the end of the observation period; hence, the survival time is said to be *right censored*. It is not known when (or, indeed, whether) a subject will experience the event, only that the subject has not experienced the event by the end of the observation period. There are other ways in which right censoring may occur. Subjects in a study may drop out during the observation period or may be lost to follow-up during the study. In some cases, they may even experience a "competing" event that makes further follow-up impossible. A good example of this is patients being followed for a cardiac event such as myocardial infarction may die from another disease or in an accident.

Unlike right censoring, an observed response is *left censored* if the corresponding event of interest has already occurred before the subject is enrolled in a study. Left censoring is common in pharmacokinetics, and it occurs in the quantification of drug concentrations when the measured concentration is below the limit of quantification of the assay. Left censoring also occurs in studies where subject recall is the measurement method. An example would be, "When did you first smoke a cigarette?"

When the event of interest is known to occur only within an interval of nonzero length, *interval censoring* is said to occur. Interval censoring occurs when the presence of a medical condition or event is evaluated during periodic exams. Interval censoring is a generalization of left and right censoring patterns. When the left endpoint is 0 and the right endpoint is a censored value, then the response is said to be left-censored.

The combination of censoring and differential follow-up creates some unusual difficulties in the analysis of such data that cannot be handled properly by standard statistical methods. Thus, a different approach called *survival analysis* or *censored survival analysis* was developed for the analysis of such data.

25.2.1.1 Classification of Right Censored Data

Right censored data can be broadly classified into singly censored and progressively censored data (13). Single censoring is of two types—one in which the study is terminated after some fixed length of time, and another in which the study is terminated after some fixed number of failures has taken place. The censored data in either case are termed singly censored data. For progressively censored data, censoring time is not identical. Thus, singly censored data are classified into Type I and Type II censored data, and progressively censored data are classified as Type III censored data (14).

Censoring Type I   Type I censoring occurs when observations are made within prespecified fixed time limits, resulting in a random number of censored observations. An example of such censoring occurs when subjects are enrolled in a study of a given duration, and the event of interest has not occurred in some of the subjects by the end of the observation period. The censoring time will be identical for all such subjects and will equal the prespecified study duration. It is also possible that some subjects will drop out of the study or be lost to follow-up and will have censored observations that are less than the study duration.

Censoring Type II   Type II censoring occurs when the number of events to be observed is prespecified and the duration of study is random. In such cases the study is continued until the prespecified number of events occur, and the data from subjects who have not had the event are censored at an identical value, but this value is not known a priori. This type of censoring is similar to Type I censoring, as the censored time is identical for all subjects who did not drop out of the study.

Type II censoring has the significant advantage that one can specify in advance how many subjects are to experience the event, and this helps to ensure that sufficient time to event observation is available to allow meaningful characterization of the time to event distribution. However, an open-ended random study period is generally impractical and this type of study is rarely seen.

Censoring Type III   Type III is differentiated from Type I and II censored data, by the censored times that are not identical, even for subjects who do not drop out of a study. This type of censoring occurs when the study is of fixed duration, and the event of interest is duration of a response that is first observed at a random time after the start of the study. As the starting time of the response is random, the censoring time for all subjects who remain enrolled at the end of the study will also be random.

In some cases, the exact time at which an event occurs is not known, but the event is known to have occurred between two recorded times. Such cases are termed *interval censored*. This type of censoring is present in the analgesic trial example presented later in this chapter. Survival time analysis is better suited than logistic analysis to the analysis of interval or right censored data.

25.2.2   Functions for Survival Time and Relationships of the Survival Functions

25.2.2.1   Distribution of Time to an Event
When analyzing survival data, summary statistics may not have the desired statistical properties, such as unbiasedness, because of possible censoring. The sample mean, for instance, is no longer an unbiased estimator of the population mean (of survival time). Thus, other methods are needed for presenting such data. An approach would be estimating the underlying true distribution. With the distribution estimated, it is then possible to estimate other quantities of interest such as median or mean. The distribution of the random variable T can be described by the usual cumulative distribution function

$$F(t) = P[T \leq t], \quad t \geq 0 \tag{25.1}$$

which is right continuous; that is, $\lim_{u \to t+} F(u) = F(t)$. When T is a survival time, $F(t)$ is the probability that a randomly selected subject from the population will have a specified event of interest before time t. Assuming T is a continuous random variable, then its density function $f(t)$, which is related to $F(t)$, is given by

$$f(t) = \frac{dF(t)}{dt}, \quad F(t) = \int_0^\infty f(u)\, du \tag{25.2}$$

It is often common to use the *survival function* given by

$$S(t) = P[T > t] = 1 - F(t) \tag{25.3}$$

where $F(t) = \lim_{u \to t} F(u)$. When T is a survival time, $S(t)$ is the probability that a randomly selected individual will survive to time t or beyond. From the above equations, it can be seen that the relationship between the density function and the survival function is given by

$$f(t) = -\frac{dS(t)}{dt} \tag{25.4}$$

The survival function $S(t)$, which takes on the value 1 at $t = 0$ (i.e., $S(0) = 1$), is a nonincreasing function over time. For a proper random variable T, $S(\infty) = 0$.

However, it is also necessary to allow the possibility that $S(\infty) > 0$. This corresponds to a situation where there is a positive probability of the event not occurring. If, for example, the event of interest is the time from response until disease relapse and the disease has a cure for some proportion of individuals in the population, then we have $S(\infty) > 0$.

25.2.2.2 Hazard Rate

A useful way of describing the distribution of time to an event is the hazard rate because it has a natural interpretation that relates to the aging of a population. Before defining the hazard rate we first define the mortality rate, which is a discrete version of the hazard rate. The mortality rate at time t, where t is generally taken to be an integer in terms of some unit of time (e.g., days, months, years), is the proportion of the population who die between times t and $t + 1$ among individuals alive at time t:

$$m(t) = P[t \le T < t+1 | T > t] \tag{25.5}$$

The hazard rate, $\lambda(t)$, is the limit of the mortality rate if the interval of time is taken to be small (rather than one unit). The instantaneous rate of failure at time t given that an individual is alive at time t is the hazard rate. $\lambda(t)$, therefore, is defined by the following equation:

$$\lambda(t) = \lim_{h \to 0} \frac{P[t \le T < t+h | T > t]}{h} \tag{25.6}$$

This can be expressed as

$$\lambda(t) = \frac{\displaystyle\lim_{h \to 0} \frac{P[t \le T < t+h]}{h}}{P[T \ge t]} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d \ln(S(t))}{dt} \tag{25.7}$$

Integrating both sides, we obtain

$$\Lambda(t) = \int_0^t \lambda(u)\, du = -\ln(S(t)) \tag{25.8}$$

where $\Lambda(t)$ is referred to as the cumulative hazard function, and $S(t)$ is the fraction surviving at time t. It is assumed that $S(0) = 1$. Hence,

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u)\, du\right) \tag{25.9}$$

There is a one-to-one relationship between the hazard rate $\lambda(t)$, $t \ge 0$ and the survival function $S(t)$, given by the formulas above. Note that the hazard rate is a probability rate, and not a probability. Thus, it is possible that a hazard rate may exceed 1 in the same fashion as a density function $f(t)$ may exceed 1.

If the hazard is constant, that is, $\lambda(t) = \lambda$ for all $t \ge 0$, then $S(t) = e^{-\lambda t}$. This distribution is the exponential distribution with hazard equal to λ. The Weibull model

is another class of distribution widely used in survival analysis, where the survival function is given by

$$S(t) = \exp(-\lambda t^a), \quad a, \lambda > 0 \tag{25.10}$$

The hazard function for the Weibull model is

$$\lambda(t) = a\lambda t^{a-1} \tag{25.11}$$

The model allows for a constant hazard ($a = 1$), increasing hazard ($a > 1$), and decreasing hazard ($0 < a < 1$).

25.3   NONLINEAR MIXED EFFECTS MODELING APPROACH TO THE ANALYSIS OF NONRANDOMLY CENSORED ORDERED CATEGORICAL LONGITUDINAL DATA FROM ANALGESIC TRIALS

Pain relief scores collected over a period of time among groups of subjects are usually compared in a clinical trial of an analgesic agent. Different subjects are randomly assigned doses of active agent or placebo when they first request it after experiencing the same painful procedure such as a third molar extraction. The data constitute short individual time series of ordered categorical pain relief scores subsequent to dosing. Since patients can elect to remedicate with an active agent if their pain relief is insufficient, nonrandom right censoring may be present. The trial is usually designed to address two questions such as: (a) Does the drug relieve pain? If so, (b) What dosage regimen should be recommended for use by a typical patient or investigated further?

The analysis of analgesic trial data are complicated by several factors:

- Repeated measurements are obtained per patient.
- The responses measured are not continuous—pain relief is often measured as an ordered categorical variable, while time for remediation is a survival variable.
- Pain relief scores are nonrandomly censored, meaning that subjects with less pain relief are more likely to seek remediation.

The nonrandom censoring creates a biased sample of patients, especially at the later time points. This is because patients sensitive to drug treatment and who experience pain relief are the ones who will not remediate. To derive the pharmacodynamic relationships and decide on the appropriate dose required to achieve adequate pain relief, only the unconditional pain relief measurements are relevant to address the question of whether the drug causes pain relief relative to placebo.

Traditional (ANOVA) analysis of analgesic clinical trials (i.e., testing the null hypothesis when comparing treatment and placebo groups) have dealt inadequately with the complexities of pain relief data collected in these studies (3, 15). When patients have required rescue medication before the end of the study, scores of unobserved subsequent pain and pain relief (PR) scores have historically been imputed according to predetermined rules such as the so-called last observation

carried forward (LOCF) imputation scheme. Evaluation of pain relief data with this imputation scheme, which has no explicit assumption, has been shown to significantly underestimate the response to treatment, to overestimate placebo-corrected drug response, and to yield a biased dose–response relationship (7). In addition, the traditional approach used in the analysis of analgesic clinical trials results in a loss of all information on the individual patient and fails to render any insight into the intersubject variability associated with the "population average" dose–response relationship. It is, therefore, difficult to construct dosing strategies.

A subject-specific random effects model for the analysis of analgesic clinical trials, which accounted for the distribution of pain relief scores, time, drug concentration, and other covariates, was developed by Sheiner (3). This approach was subsequently used for the development of a model for ketorolac analgesia (5). A slight modification of the approach was introduced by Liu and Sambol (4), and it involved the use of a model-independent method (empirical convolution) to generate effect site concentrations. The effect site concentration is the concentration of drug at the site of action or biophase. This nonlinear mixed effects methodology is not limited to analgesic studies but is applicable wherever the outcome is measured as a categorical variable.

In the subsequent sections we present the nonlinear mixed effects model approach for analyzing analgesic data and apply it to simulated analgesic study data.

25.3.1 Methodology

The approach involves a semimechanistic or mechanistic model that describes the joint probability of the time of remediation and the pain relief score (which is related to plasma drug concentrations). This joint probability can be written as the product of the conditional probability of the time of remediation, given the level of pain relief and the probability of the pain relief score. First, a population pharmacokinetic (PK) model is developed using the nonlinear mixed effects modeling approach (16–19) (see also Chapters 10 and 14 of this book). With this approach both population (average) and random (inter- and intraindividual) effects parameters are estimated. When the PK model is linked to an effect (pharmacodynamic (PD) model), the effect site concentration (C_e) as defined by Sheiner et al. (20) can be obtained. The effect site concentration is useful in linking dose to pain relief and subsequently to the decision to remediate.

To model the distribution of pain relief scores and remediation times, subject-specific random effect models are developed. Let the vector of pain relief scores for an individual be $Y = (Y_1, Y_2, \ldots, Y_N)$. At time t the pain relief score is denoted by Y_t and the time at which an individual remediates is denoted by the variable T. The PD model parameter estimates are obtained by maximum likelihood, which estimates the most probable model parameter values for the observed data. $P[T, Y]$ denotes the likelihood of an individual's data, and it is expressed by the following equation:

$$P[T, Y] = \int P[T, Y | \eta] \cdot P[Y | \eta] \cdot P[\eta] \, d\eta \qquad (25.12)$$

where η is a vector of subject-specific random effects, assumed to be (multivariate) normally distributed with a mean of zero and variance Ω. The likelihood can

be factored out in two terms: one related to pain relief ($P[Y \mid \eta]$), and the other related to the remedication behavior conditional on pain relief ($P[T \mid Y, \eta]$). In the subsequent sections, models for subject-specific distributions $P[Y \mid \eta]$ and $P[T \mid Y, \eta]$ are discussed.

25.3.1.1 *Pain Relief Model P[Y | η]*
Pain relief is a categorical variable that can take a value of 0 (no pain relief), 1 (a little pain relief), 2 (some pain relief), 3 (a lot of pain relief), or 4 (complete pain relief). The log-odds that Y_t is greater than or equal to the score m ($m = 1, \ldots, 4$) is given by

$$g\left(P[Y_t \geq m | \eta_Y]\right) = f_p(m, t) + f_d(C_e) + \eta_Y \qquad (25.13)$$

where $g(x)$ denotes the logit transform of the probability of an event that ensures probability values between 0 and 1, f_p is the function describing the time course of the placebo effect, f_d is the function describing the drug effect, and η_Y is a random individual effect determining individual sensitivity. The η_Y are assumed to be normally distributed with standard deviation ω_Y.

The probability distribution of pain relief scores is given by the inverse logit of $g(x)$:

$$P[Y_t \geq m | \eta_Y] = \frac{\exp\left(f_p(m, t) + f_d(C_e) + \eta_Y\right)}{1 + \exp\left(f_p(m, t) + f_d(C_e) + \eta_Y\right)} \qquad (25.14)$$

with the placebo effect given by

$$f_p(m, t) = \sum_{k=0}^{m} \beta_k + A\left[e^{\alpha t} - e^{\gamma t}\right] \qquad (25.15)$$

where α and γ are first-order rate constants of the offset and onset of the placebo effect. A is a scaling parameter that determines the size of the placebo effect, and β_k specifies the baseline set of probabilities of the various degrees of pain relief. There is no estimate for β_0 in the model because $P[Y_t \geq 0] = 1$. The probability that $P[Y_t = m]$ is then equal to the difference in probabilities of two subsequent pain relief scores, that is, $P[Y_t \geq m] - P[Y_t \geq (m-1)]$.

Models of varying complexities can be used to describe the placebo effect. Models for drug effect can be semimechanistic (i.e., link model) (20) or mechanistic (i.e., indirect response model) (21). A semimechanistic drug effect model can be expressed as

$$f_d(m, C_e, t) = \frac{E_{\max} \cdot C_e}{EC_{e50} + C_e} \qquad (25.16)$$

where $E_{\max}$ is maximum drug effect, and EC_{e50} is the effect compartment drug concentration at 50% of the maximal drug effect. C_e denotes effect site concentration and is given by

$$C_e(t) = \int_0^t k_{e0} C_p(u) e^{-k_{e0}(t-u)} du \qquad (25.17)$$

where k_{e0} is the first-order rate constant that characterizes the delay between plasma and effect site concentrations. The population average PK parameters derived from the PK analysis are used to calculate $C_p(t)$, the concentration of the drug in the plasma at time t. Assuming that pain relief scores within an individual at distinct times are independent, the vector of pain relief scores $P[Y \mid \eta_Y]$ is given by

$$P[Y|\eta_Y] = \prod P[Y_t|\eta] \tag{25.18}$$

25.3.1.2 Model for Remediation P[T | Y,η]

The time to remediation can be viewed as a survival variable (22). By definition, a survival function, $S(t)$, is the probability that a person remains in the study (does not remediate) up to time t and is given by

$$P[T > t|Y, \eta] = S(t) = \exp\left(-\int_0^t \lambda(u)\,du\right) \tag{25.19}$$

where $\lambda(t)$ is the hazard function. The hazard function can be interpreted as an instantaneous risk, in that $\lambda(t)\,\delta t$ is the probability that a subject remediates in the next small interval of time δt, given that he/she has not remediated. A constant hazard over a fixed interval of time indicates that a constant proportion of patients who are still in the study are expected to remediate. A constant hazard function implies an exponential distribution of remediation with mean $1/\lambda$ and hence a Poisson process. Several models can be used to evaluate the hazard function, and the model that best describes the data is used to describe remediation (3, 5, 7, 8).

A model that allows the baseline hazard to change linearly over time but remain constant over a time interval δt is expressed as

$$\lambda(t, Y_t, \eta) = \lambda_m [1 + FH(t-1)_+] \exp(\eta_T) \tag{25.20}$$

where λ_m is the baseline hazard rate, FH is the fractional change in λ_m with time, $(t-1)_+ = t-1$ for $t > 1$ and zero otherwise, and η_T is a random individual effect. The assumption implicit in this model is that $P[T \mid Y, \eta]$ depends only on η_T and the observable elements of Y. $P[T > t \mid Y, \eta]$ is set to 1 for the time points before remediation is allowed. This time can be either 1, 2, 3, or 4 h, depending on the study design. However, the hazard is allowed to accumulate according to Eq. (25.9) independent of the first time remediation is allowed. It follows from Eq. (25.8) that the probability that an individual will remediate at time t, given the individual is still in a study at the previous observation time $t-1$, is given by

$$P[T = t|T \geq t] = 1 - \frac{S(t)}{S(t-1)} = 1 - \exp\left(-\int_{t-1}^t \lambda(t)\,dt\right) \tag{25.21}$$

25.3.2 Estimation and Inference

The Laplacian estimation method as implemented in the NONMEM program (a program the authors use for nonlinear mixed effects modeling) is used to provide

maximum likelihood estimates of model parameters (23). Assuming the individuals to be independent, the likelihood L for all the data from N subjects can be specified by the product of the probability of each subject's data:

$$L = \prod_{i=1}^{N} \int P[T, Y | \eta] P[\eta] d\eta$$
$$L = \prod_{i=1}^{N} \int P[T | Y, \eta] P[Y | \eta] P[\eta] d\eta \tag{25.22}$$

In order to simplify the calculations, it can be assumed that η_Y and η_T are independent, that is, $\mathrm{Cov}(\eta_Y, \eta_T) = 0$. The implication of this is that pain relief data can be fitted separately from the remedication data by independent maximization of the following likelihoods:

$$L = \prod_{i=1}^{N} \int P(Y | \eta_Y) P(\eta_Y) d\eta_Y \tag{25.23}$$

$$L = \prod_{i=1}^{N} \int P(T | Y, \eta_Y) P(Y | \eta_Y) P(\eta_Y) d\eta_Y \tag{25.24}$$

Model selection is based on the likelihood ratio test with $p < 0.001$ and diagnostic plots. The difference in minus twice the log of the likelihood ($-2LL$) between a full and a reduced model is asymptotically χ^2 distributed with degrees of freedom equal to the difference in the number of parameters between two models. At $p < 0.001$, a decrease of more than 6.6 in $-2LL$ is significant. Asymptotic standard errors are obtained from the asymptotic covariance matrix. Alternatively, confidence intervals on parameters can be computed for this very nonlinear situation from the likelihood profile plot (24).

25.3.3 Prediction

Once the population model has been developed, interesting population statistics (time to onset of effect, percent of patients at peak effect) can be computed by means of Monte Carlo integration with respect to η (see Refs. 1, 4, 6, and 8). However, this is not addressed in the example in Section 25.4. By simulating η values from the estimated distribution, response profiles for individual subjects are generated. The population mean probability of having a certain pain relief score and the population mean expected pain relief score at a specific time and dose can be computed from these profiles. Virtual patient populations can be simulated for all doses. The goodness of fit of the model to the data can be judged by comparing model-generated simulations of the probability that pain relief is greater than or equal to m conditional on the remedication times, $P[Y \geq m \mid T \geq t]$, and model-generated estimates of the probability that a patient will remedicate at time t given the patient is still in the study up to that time point, $P[T = t \mid T \geq t]$, with data-derived estimates of these probabilities (see Refs. 1, 4, 6, and 8). The latter is not covered in the simulated example in Section 25.4.

A simulated example is presented next to illustrate these concepts. The parameters used for the simulation were modified from the example published by Ette et al. (8).

25.4 APPLICATION

In this section we present an example of the aforementioned methodology by simulating and analyzing data for a trial in which the efficacy of an analgesic drug was investigated. A description of the study design is provided, followed by a description of the pharmacokinetic (PK) and pharmacodynamic (PD) characteristics of the drug, along with the S-Plus and NONMEM codes used to simulate the trial. Finally, the parameter estimation details for the pain and remediation models are provided. The focus here is on the PD aspects of the drug, and therefore the estimation of PK parameters is not described. A sequential approach is taken in the estimation of the PD parameters in the pain and remediation models, by assuming the PK parameters for each individual have been previously estimated. In this case, the PK parameter estimates were taken to be identical to that used to simulate the data.

25.4.1 Study Design

Data were simulated based on a three-arm parallel group design (200 subjects/arm), with one placebo, and two active arms (0.5 and 1 mg). Subjects received a single dose of the study drug at the onset of a pain event, following which their pain relief scores were recorded at 0.25, 0.5, 0.75, 1, 1.5, and 2 hours. The pain relief was scored on a 5-point scale (0, none; 1, a little; 2, medium; 3, a lot; and 4, complete). It was assumed that the intensity of pain prior to receiving the drug was identical (or balanced) across the three arms of the study. Subjects were allowed to opt for rescue medication at any time and were remedicated with a drug that is the standard of care.

25.4.2 Population Pharmacokinetic and Pharmacodynamic Models

The pharmacokinetics of the analgesic drug was described by a two-compartment model, with first-order absorption and absorption lag time. The population parameters of this model are given in Table 25.1. The value of a given parameter P_k, for subject i, is given by

$$P_{ki} = P_k \exp(\eta_{ki}) \qquad (25.25)$$

where P_k is the population average value, and η_{ki} is a normally distributed random variable with mean zero and standard deviation given by the interindividual variability for the parameter. The parameters were assumed to be identical and

TABLE 25.1 Population Pharmacokinetic Model Parameters

Parameter	Symbol	Population Average	Interindividual Variability
Apparent clearance (L/h)	CL/F	2	30%
Apparent volume (central compartment) (L)	VC/F	10	30%
Apparent intercompartmental clearance (L/h)	Q/F	1	0%
Apparent volume (peripheral compartment) (L)	VP/F	20	0%
Absorption rate constant (h^{-1})	KA	2	70%
Absorption lag time (h)	TLAG	0.1	100%
Relative bioavailability	FBIO	1	70%

TABLE 25.2   Population Pain Relief Model Parameters

Parameter Description[a]	Parameter Symbol	Value
Pain Relief Model		
Baseline placebo effect ($m = 1$)	β_1	−2.5
Baseline additional placebo effect ($m = 2$)	β_2	−2.0
Baseline additional placebo effect ($m = 3$)	β_3	−1.5
Baseline additional placebo effect ($m = 4$)	β_4	−1.0
Magnitude of placebo effect	A	3
Placebo effect onset rate (h^{-1})	α	0.0
Placebo effect decay rate (h^{-1})	γ	1
Effect compartment elimination rate (h^{-1})	k_{e0}	0.5
Maximal effect of drug	E_{max}	10
Concentration to achieve 50% E_{max} (ng/mL)	EC_{e50}	40
Interindividual variability in pain relief	ω_{PR}	1
Remedication Model		
Hazard ($m = 0$)	λ_0	0.5
Hazard ($m = 1$)	λ_1	0.005
Hazard ($m = 2$)	λ_2	0.005
Hazard ($m = 3$)	λ_3	0.001
Hazard ($m = 4$)	λ_4	0.0001
Interindividual variability in hazard	ω_{HZ}	0

[a] m = pain relief score.

independently distributed. Correlations between apparent clearances and apparent volumes due to differences in bioavailability between subjects are accounted for by the relative bioavailability parameter, the population average value of which was fixed to 1 by definition.

The pain relief model described in Section 25.3.1.1 was implemented with the parameter values presented in Table 25.2. In this example, it is assumed that the placebo effect decreases monotonically with time, by setting the placebo onset rate (α) to zero. The negative values of the baseline placebo effect values (β's) indicates that the probability of a high pain relief response is less than that of a lower pain relief score.

The remedication model described in Section 25.3.1.2 was implemented with the parameter values presented in Table 25.2. The hazard parameter (λ) values indicate that the probability of remedication was highest for subjects who had no pain relief ($m = 0$), but decreased sharply for subjects who had even a small degree of relief. The hazard of remedication for subjects who had complete pain relief ($m = 4$) is fixed to zero, indicating that there was no probability that these subjects would seek remedication.

25.4.3   Model Implementation

This section describes the S-Plus and NONMEM codes used to simulate pain relief scores and remedication events, and the estimation PD parameters from the simulated data.

The S-Plus code in Appendix 25.1 was used to generate comma separated variables (CSV) NONMEM data set in accordance with the study design described earlier. This code can readily be modified to produce data sets for alternative study designs that differ in sample size, number of study arms, and pain relief score observations.

The NONMEM control file provided in Appendix 25.2 (Model Sim) was used to simulate pain relief scores and remediation events according to the population PK/PD models described previously. The NONMEM control file provided in Appendix 25.5 (Model PR+RM) was used to estimate the parameters in the pain relief and remediation models, from the simulated pain relief data and PK parameter estimates for each individual. Separate NONMEM control files are provided for simulation and estimation, because the dependent variables are not identical for simulation and estimation. The dependent variables for simulation are pain relief scores and remediation events, whereas the dependent variables for estimation are the probabilities of observing these scores and events. The pain relief scores and remediation events are simulated by utilizing a uniform random variable to assign a pain relief score or remediation event, given the respective simulated probabilities.

The PK parameters of individual subjects are not estimated in the example provided, as the focus is on the estimation of the pain and remediation model parameters. As described earlier, it is assumed that the individual PK parameter estimates are known by providing the simulated individual PK parameter values as part of the data input to Model PR+RM. However, before the parameters were estimated, the simulated data needed to be processed to remove all observations within a subject that followed a remediation event. This processing generates a nonrandomly censored data set, as subjects with lower pain relief scores are more likely to drop out of the study. The S-Plus code to perform this processing of the simulated data is provided in Appendix 25.3. This script generates a CSV file that serves as input to the NONMEM control file in Appendix 25.5. The potential for bias resulting from the nonrandomly censored data was examined by comparing the pain relief model parameter estimates obtained with Model PR+RM versus that obtained by Model PR, which is identical to Model PR+RM except that it does not have a remediation model.

25.4.4   Results

The simulated pain relief scores and remediation events for one trial are shown in Figure 25.1, which was created using the S-Plus code provided in Appendix 25.4. As expected, the simulated data indicated that the number of subjects remaining in the trial decreased with time, and that the decrease was greatest for the placebo group. The stacked bars also indicate that the level of pain relief tended to increase with dose and time.

The parameter estimates obtained with Models PR+RM and PR are presented in Table 25.3, along with the relative standard error (RSE) and the bias in the estimates. In the interest of generality, the estimation model retained several parameters (e.g., α and ω_{HZ}) that were set to zero in the simulation model. All the nonzero parameter values in the simulation model were estimated except λ_3 and λ_4, which could not be estimated with this data set, and were fixed to their true values. All

FIGURE 25.1 Stacked barplot of pain relief scores versus time by dose. The effect of remedication is seen by the decrease in number of subjects remaining in the study as time progresses.

the parameters were estimated with relatively small uncertainty, except for k_{e0} and EC_{e50}. The difficulty in estimating k_{e0} (and this EC_{e50}) might be due to the short observation window of 0.25–2 hours. The RSE was less than the bias for all of the parameters, suggesting that uncertainty estimated by NONMEM is realistic.

A graphical assessment of the goodness of fit of Model PR+RM is presented in Figure 25.2. In the figure, model-estimated probability of pain relief is compared with the proportion of subjects exhibiting pain relief. Good agreement between the data and model-predicted probability was achieved by the estimated parameters. The plot for Model PR is comparable with that of Model PR+RM. The S-Plus code to create the plot in Figure 25.2 from the NONMEM table output is presented in Appendix 25.6.

25.5 OTHER METHODS FOR ANALYZING ORDERED CATEGORICAL DATA

Jönsson (25) showed from a simulation study that the use of the standard mixed effects modeling approach may produce biased parameter estimates when ordered categorical data with a skewed distribution are analyzed using the Laplacian method. Increasing interindividual variability and skewness in the distribution of the data increase the bias associated with the estimation of those parameters. The conse-

TABLE 25.3   Parameter Estimates: Population Pharmacodynamic Model

Parameter Symbol	Model PR + RM			Model PR		
	Estimate	RSE[a] [%]	Bias[b] [%]	Estimate	RSE[a] [%]	Bias[b] [%]
Pain Relief Model						
β_1	−2.91	11.9	16.4	−2.91	11.9	16.4
β_2	−2.02	3.2	1.0	−2.02	3.2	1.0
β_3	−1.52	4.7	1.3	−1.52	4.7	1.3
β_4	−0.999	6.6	−0.1	−0.999	6.6	−0.1
A	3	9.0	0.0	3	9.0	0.0
α	0		0.0	0		0.0
γ	1.64	25.5	64.0	1.64	25.4	64.0
k_{e0}	0.248	76.6	−50.4	0.248	76.6	−50.4
E_{max}	10.3	8.2	3.0	10.3	8.2	3.0
EC_{e50}	22.8	73.7	−43.0	22.8	73.7	−43.0
ω_{PR}	0.83	12.0	−17.0	0.83	12.0	−17.0
Remediation Model						
λ_0	0.147	7.5	−70.6	NA	NA	NA
λ_1	0.00213	70.9	−57.4	NA	NA	NA
λ_2	0.0036	70.6	−28.0	NA	NA	NA
λ_3	0.001 (fixed)	NA	NA	NA	NA	NA
λ_4	0.0001 (fixed)	NA	NA	NA	NA	NA
ω_{HZ}	0 (fixed)	NA	NA	NA	NA	NA

[a] RSE = | 100*(standard error)/estimate |.
[b] Bias = 100*(estimate − trueValue)/trueValue.

quence of this is the overestimation of the frequency of rare events when simulation is performed with biased parameter values. To deal with this, two methods have been proposed to model categorical responses as a two-step process (26, 27). The first step involves modeling the incidence of response or the probability of being a responder and the second step involves modeling the severity of response. Olsen and Schafer (27) proposed modeling the response as a semicontinuous variable. They use a logistic model to model the incidence of effect and a continuous model for the severity of effect. In both steps, intersubject variability was included. Thus, in the first step, the polychotomous data are reduced to binary data by modeling the presence or absence of an effect (without considering the severity of the effect), thereby decreasing the skewness of response. The result of this is a reduction in the bias in parameter estimation.

On the other hand, Kowalski et al. (26) modeled the first step similar to that proposed by Jonssen but only considered one observation per individual without modeling interindividual variability. The severity of response (given being a responder) was modeled using the standard mixed effects modeling approach in the second step. The noninclusion of intersubject variability in the first step enabled them to avoid bias in parameter estimation, similar to the Olsen–Schafer model. Thus, skewness of the response distribution was reduced by only including responders in the second step. Dose escalation studies and time-varying covariates cannot be

FIGURE 25.2 Proportion and probability of pain relief scores versus time by dose. The numbered symbols represent calculated proportion of subjects with pain relief greater than or equal to the number ($m \geq 1, 2, 3,$ or 4), and the lines represent the corresponding model predicted probabilities (Model PR+RM).

handled appropriately using this method since the first step only considers one observation per patient. Another drawback of the method is the fact that in the first step the parameter estimates will be dependent on the number of observations per subject. Thus, there is a tendency of increasing probability of a nonzero event with increasing number of observations. The possibility of making extrapolations based on the model is limited because of the data dependency of the model parameter values. The back-step method (BSM) has been proposed (28) as an alternative to changing the model to handle biased estimates—that is, modifying the estimation procedure itself. The method is an iterative approach that involves searching for the unbiased parameter estimates, which upon simulation generate data that mimic the original data.

25.6 SUMMARY

An overview of categorical data and methods used to analyze such data together with survival data, censoring, and survival functions is presented. Ordered categorical data can be analyzed with marginal and/or conditional models depending on the objective of the analysis. Marginal models are empirical models that characterize population-averaged effects, and generalized estimation equations are an example of such models. Conditional models, on the other hand, characterize subject-specific effects as well as population-averaged effects. The use of conditional models, such

as a nonlinear mixed effects model, to analyze ordered categorical data is discussed with particular application to analgesic trials. The nonlinear mixed effects modeling approach involves a model that links pain relief and the decision to remediate to pharmacokinetics and dosage, within a semimechanistic pharmacological framework. The advantages of this methodology over the traditional ANOVA approach is also presented. A simulated example is presented with NONMEM and S-Plus codes that can be used to simulate and analyze pain relief with or without remedication. Other modifications of the mixed effects modeling approaches, developed to deal with bias in parameter estimates when there is skewness in the distribution of ordered categorical response data, are discussed. The application of the nonlinear mixed effects methodology described in this chapter is not limited to analgesic trial data, but is applicable wherever the outcome is measured as a categorical variable.

REFERENCES

1. C. A. Knibbe, K. P. Zuideveld, J. DeJongh, P. F. Kuks, L. P. Aarts, and M. Danhof, Population pharmacokinetic and pharmacodynamic modeling of propofol for long-term sedation in critically ill patients: a comparison between propofol 6% and propofol 1%. *Clin Pharmacol Ther* **72**:670–684 (2002).

2. P.-H. Zingmark, M. Ekblom, T. Odergren, T. Ashwood, P. Lyden, M. O. Karlsson, and E. N. Jonsson, Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol* **56**:173–183 (2003).

3. L. B. Sheiner, A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther* **56**:309–322 (1994).

4. C. Y. Liu and N. C. Sambol, Pharmacodynamic analysis of analgesic trials using empirical methods. *Pharm Res* **12**:438–445 (1995).

5. J. W. Mandema and D. R. Stanski, Population pharmacodynamic model for ketorolac analgesia. *Clin Pharmacol Ther* **60**:619–635 (1996).

6. L. B. Sheiner, S. L. Beal, and A. Dunne, Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *J Am Stat Assoc* **92**:1235–1255 (1997).

7. J. W. Mandema, Population pharmacokinetics/pharmacodynamics of analgesics: theory and applications, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, M. Danhof, M. Gex-Fabry, U. A. Gundert-Remy, M. O. Karlsson, F. Mentre, P. L. Morselli, F. Rombout, M. Rowland, J. L. Steimer, and S. Vozeh (Eds.). Commission of the European Communities, European Cooperation in the Field of Scientific and Technical Research, Brussels, 1997, pp. 74–82.

8. E. I. Ette, P. Lockwood, R. Miller, and J. W. Mandema, Analysis of analgesic trials, in *Applied Statistics in the Pharmaceutical Industry Using S-Plus* S. Millard and A. Krause (Eds.). Springer-Verlag, New York, 2001, pp. 237–266.

9. A. Agresti, *Categorical Data Analysis*. Wiley, Hooken, NJ, 2002.

10. A. Agresti, Modelling ordered categorical data: recent advances and future challenges. *Stat Med* **18**:2191–2207 (1999).

11. P. McCullagh, Regression models for ordinal data. *J R Stat Soc B* **42**:109–142 (1980).

12. B. Peterson and F. E. Harrell, Partial proportional odds models for ordinal response variables. *Appl Stat* **39**:205–217 (1990).

13. A. C. Cohen, Jr., Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics* **7**:579–588 (1965).

14. E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis*. Wiley, Hoboken, NJ, 2003.

15. E. M. Laska, C. Siegel, and A. Sunshine, Onset and duration: measurement and analysis. *Clin Pharmacol Ther* **49**:1–5 (1991).

16. A. Racine-Poon and J. Wakefield, Statistical methods in population pharmacokinetic modeling. *Stat Methods Med Res* **7**:63–84 (1998).

17. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

18. E. I. Ette, P. Williams, H. Sun, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–54 (2001).

19. E. I. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M.-J. Liu, and E. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).

20. L. B. Sheiner, D. R. Stanski, S. Vozeh, and R. D. Miller, Simultaneous modeling of pharmacokinetics modeling of pharmacokinetic and pharmacodynamics: application to D-tubocurarine. *Clin Pharmacol Ther* **25**:358–371 (1979).

21. N. I. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

22. P. McCullagh and J. A. Nelder, *Generalized Linear Models*. Chapman & Hall, New York, 1989.

23. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. Globomax, Hanover, MD, 1988–1999.

24. D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and Its Application*. Wiley, Hoboken, NJ, 1988.

25. S. Jonsson, *Estimation of Dosing Strategies for Individualization*. Dissertation, Acta Universitatis Upsaliensis, Uppsala, Sweden, 2004. Available at `http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4255`.

26. K. G. Kowalski, L. McFadyen, M. M. Hutmatcher, B. Frame, and R. Miller, A two-part model for longitudinal adverse event severity data. *J Pharmacokinet Biopharm* **30**:315–336 (2003).

27. M. K. Olsen and J. L. Schafer, A two-part random-effects model for semicontinuous longitudinal data. *J Am Stat Assoc* **96**:730–745 (2001).

28. M. C. Kjellsson, S. Jonsson, and M. O. Karlson, The back-step method for obtaining unbiased population parameter estimates for ordered categorical data. *AAPS J* **6**(3): E1–E10 (2004).

APPENDIX 25.1   S-PLUS CODE TO GENERATE NONMEM DATA SET

```
######################### Generate NONMEM data set #########################

##### Specify parameters for clinical trial
nArm <- 3
nSubjArm <- 200
dose.v <- c(0, 500, 1000) # ug
pdObsTime.v <- c(0, 15, 30, 45, 60, 90, 120)/60 # [hr]
```

```
##### Calculate clin trial design parameters
nSubj <- nArm*nSubjArm
subj.v <- 1:nSubj
nDose <- length(dose.v)
nPDObs <- length(pdObsTime.v)


# Vars needed for NONMEM
# ID TIME MDV PRLF=DV QUIT PTIM DOSE
# ID . . . NONMEM ID
# TIME . . . Dose or Observation Time [hr]
# PRLF . . . Pain relief score
# QUIT . . . Remedication indicator
# (0=subject stays in trial, 1=subject remedicates and quits trial)
# PTIM . . . Time of previous observation
# DOSE . . . Dose [ug]


##### Create PD observation records
pdObs.tmp <- data.frame(ID=subj.v,
  TIME=rep(NA, nSubj),
  MDV=rep(NA, nSubj),
  PRLF=rep(NA, nSubj),
  QUIT=rep(NA, nSubj),
  PTIM=rep(NA, nSubj),
  DOSE=rep(dose.v, each=nSubjArm))


##### Create PD observation records
pdObs <- pdObs.tmp
for (iObs in 2:nPDObs){
  pdObs <- rbind(pdObs, pdObs.tmp)
}
#
pdObs$TIME <- rep(pdObsTime.v, each=nSubj)
pdObs$MDV <- rep(0, nrow(pdObs))
pdObs$PRLF <- rep(NA, nrow(pdObs))
pdObs$QUIT <- rep(NA, nrow(pdObs))


TF.v <- pdObs$TIME==0
pdObs$MDV[TF.v] <- rep(1,sum(TF.v))


pdObs <- pdObs[order(pdObs$ID, pdObs$TIME), ]


##### specify "previous time"
pTime <- pdObs$TIME[pdObs$ID==1]
pTime <- c(0, pTime[-length(pTime)])
pdObs$PTIM <- rep(pTime, nSubj)


pdObs <- pdObs[order(pdObs$ID, pdObs$TIME, -pdObs$MDV), ]
```

```
exportDir <- "../"
fileName <- paste(exportDir, "analgesicTemplate.csv", sep="")


##### Define function to export NONMEM datasets as CSV file
z.export.csv <- function(data.df, fileName, MissingVal=".", AddHash=F){
  if(AddHash){
    names(data.df)[1] <- paste("#", names(data.df)[1])
  }
  cat(names(data.df), sep=", ", file=fileName)
  cat("\n",file=fileName, append=T)
  write.table(data.df, file=fileName, sep = ",", na = MissingVal, dimnames.
  write = F, append=T)
}
z.export.csv(pdObs, fileName, MissingVal=".", AddHash=T)
```

APPENDIX 25.2 NONMEM CODE TO SIMULATE CLINICAL TRIAL

```
$PROB Analgesic Pain Model for nonrandomly censored data


$INPUT ID=L1 TIME MDV PRLF=DV QUIT PTIM DOSE
;
; ID   = subject ID number
; TIME = time of dose or observation
; EVID = event ID (0=obs, 1=dose)
; PRLF = Ordinal pain relief score (0=no relief thru 4=full relief)
; TQT  = Time to remediation
; QUIT = Indicator of remediation (0=stay in study; 1=quit study)
; PTIM = Time of previous pain observation
; DOSE = nominal dose amount


$DATA ../analgesicTemplate.csv IGNORE=#


$THETA
; PK Model Parameters
2.0 FIX ; CL/F [L/h]
10.0 FIX ; VC/F [L]
1.0 FIX ; Q/F [L/h]
20.0 FIX ; VP/F [L]
2.0 FIX ; KA [1/h]
0.1 FIX ; ALAG [h]
1 FIX ; FBIO [1]
;
; Pain Model Parameters
```

```
0.5 ; KE0 [1/h]
-2.5 ; BT1
-2 ; BT2
-1.5 ; BT3
-1 ; BT4
40 ; EC50 [ng/mL]
10 ; EMAX
0 FIX ; ALPH
1 ; GAMM
3 ; AA
;
; Remedication Model Parameters
0.5 ; LAM0
0.005 ; LAM1
0.005 ; LAM2
0.001 ; LAM3
0.0001 ; LAM4

$OMEGA
; PK Model Parameters
0.09 ; CL/F
0.09 ; VC/F
0 FIX ; Q/F
0 FIX ; VP/F
0.49 ; KA
1.0 ; ALAG
0.49 ; FBIO
;
; Pain and Remedication Model Parameters
1 ; Pain
0.0 FIX ; Remediation

$ABBREVIATED DERIV2=NOCOMMON

$PRED
; Specify PK Model Parameters
CL = THETA(1)*EXP(ETA(1))
VC = THETA(2)*EXP(ETA(2))
Q = THETA(3)*EXP(ETA(3))
VP = THETA(4)*EXP(ETA(4))
KA = THETA(5)*EXP(ETA(5))
ALAG = THETA(6)*EXP(ETA(6))
FBIO = THETA(7)*EXP(ETA(7))

; Specify Pain Model Parameters
KE0 = THETA(8)
BT1 = THETA(9)
BT2 = THETA(10)
```

```
BT3 = THETA(11)
BT4 = THETA(12)
EC50 = THETA(13)
EMAX = THETA(14)
ALPH = THETA(15)
GAMM = THETA(16)
AA = THETA(17)
ZPAN = ETA(8)
;
; Specify Remedication Model Parameters
LAM0 = THETA(18)
LAM1 = THETA(19)
LAM2 = THETA(20)
LAM3 = THETA(21)
LAM4 = THETA(22)
ZRMD = ETA(9)


; Calculate concentration in effect compartment

  K20 = CL/VC
  K23 = Q/VC
  K32 = Q/VP
  BET1 = K23+K32+K20
  BET2 = SQRT(BET1**2 - 4*K32*K20)
  BETA = 0.5*(BET1 - BET2)
  ALFA = K32*K20/BETA

  BSL = KE0*KA*DOSE*FBIO/VC
  M1 = ALFA - KA
  M2 = -M1
  Q1 = BETA - KA
  Q2 = -Q1
  R1 = KE0 - KA
  R2 = -R1
  S1 = BETA - ALFA
  S2 = -S1
  T1 = KE0 - ALFA
  T2 = -T1
  U1 = KE0 - BETA
  U2 = -U1
  Z1 = K32 - KA
  Z2 = K32 - ALFA
  Z3 = K32 - BETA
  Z4 = K32 - KE0

  TIM2 = TIME-ALAG
  E1 = EXP(-KA*TIM2)
  E2 = EXP(-ALFA*TIM2)
```

```
  E3 = EXP(-BETA*TIM2)
  E4 = EXP(-KE0*TIM2)


  CE1 = Z1*E1/(M1*Q1*R1)
  CE2 = Z2*E2/(M2*S1*T1)
  CE3 = Z3*E3/(Q2*S2*U1)
  CE4 = Z4*E4/(R2*T2*U2)
IF (TIME .LE. ALAG) THEN
  CE = 0.0
ELSE
  CE = BSL*(CE1 + CE2 + CE3 + CE4)
ENDIF

; Specify placebo effect for each cumulative probability
PEFF = EXP(-ALPH*TIME) - EXP(-GAMM*TIME)
PEFF1 = BT1 + AA*PEFF
PEFF2 = PEFF1 + BT2
PEFF3 = PEFF2 + BT3
PEFF4 = PEFF3 + BT4

; Specify drug effect
DEFF = EMAX * CE/(EC50 + CE)

; Logits for cummulative probabilities
  LGT1 = PEFF1 + DEFF + ETA(8)
  LGT2 = PEFF2 + DEFF + ETA(8)
  LGT3 = PEFF3 + DEFF + ETA(8)
  LGT4 = PEFF4 + DEFF + ETA(8)

; Exponentiate logit
C1 = EXP(LGT1)
C2 = EXP(LGT2)
C3 = EXP(LGT3)
C4 = EXP(LGT4)

; Calculate cumulative probability of response j
  P1 = C1/(1+C1) ; P(Y<=1|X)
  P2 = C2/(1+C2) ; P(Y<=2|X)
  P3 = C3/(1+C3) ; P(Y<=3|X)
  P4 = C4/(1+C4) ; P(Y==4|X)

; Likelihood (Yj), by pain relief (j = 0 to 4)
  Y0 = 1 - P1
  Y1 = P1 - P2
  Y2 = P2 - P3
  Y3 = P3 - P4
  Y4 = P4
```

```
; If PREDPP is being called for simulation, then...
;...generate uniform random number (source #2), and
;...call this random number UNIF1
;...use UNIF1 to assign the level of pain relief
;...NOTE: PRLF is named PRLS for simulation
IF(ICALL .EQ. 4) THEN
  CALL RANDOM(2, R)
  UNIF1 = R
  PRLF = 0
  IF (P1 .GT. UNIF1) PRLF=1
  IF (P2 .GT. UNIF1) PRLF=2
  IF (P3 .GT. UNIF1) PRLF=3
  IF (P4 .GT. UNIF1) PRLF=4
  PRLS = PRLF
ENDIF


  IND0=0
  IND1=0
  IND2=0
  IND3=0
  IND4=0


IF (PRLF .EQ. 0) IND0=1
IF (PRLF .EQ. 1) IND1=1
IF (PRLF .EQ. 2) IND2=1
IF (PRLF .EQ. 3) IND3=1
IF (PRLF .EQ. 4) IND4=1


; Calculate likelihood of pain score
YP = Y0*IND0+Y1*IND1+Y2*IND2+Y3*IND3+Y4*IND4
; Calculate likelihood of remedication
LAMM = LAM0
IF(PRLF .EQ. 1) LAMM = LAM1
IF(PRLF .EQ. 2) LAMM = LAM2
IF(PRLF .EQ. 3) LAMM = LAM3
IF(PRLF .EQ. 4) LAMM = LAM4
LAMM = LAMM + ETA(9)


; Probability that subject has not remedicated upto time=TIME
  YR0 = EXP(-LAMM*TIME)
;
; Probability that subject will remedicate at time=TIME, given that
; the subject has not remedicated upto time=PTIM
  ETIM = TIME - PTIM
  YR10 = 1 - EXP(-LAMM*ETIM)
  YR11 = EXP(-LAMM*PTIM)
  YR1 = YR10*YR11
```

```
; If PREDPP is being called for simulation, then . . .
; . . . generate uniform random number (source #2), and
; . . . call this random number UNIF2
; . . . use UNIF2 to determine whether subject quit
IF(ICALL .EQ. 4) THEN
  CALL RANDOM(2, R)
  UNIF2 = R
  QUIT = 0
  IF (YR1 .GT. UNIF2) QUIT=1
ENDIF


; Get simulation iteration number
ISIM = 0
IF (ICALL .EQ. 4) ISIM = IREP


;
; Likelihood for remedication model
YR = YR0*(1-QUIT) + YR1*QUIT


Y = YP*YR


$SIM (55555) (54321 UNIFORM) ONLY SUB=1
$TABLE NOPRINT ONEHEADER NOAPPEND FILE=analgesicSim.tab
  ISIM ID TIME MDV PRLS QUIT PTIM DOSE
  CL VC Q VP KA ALAG FBIO
```

APPENDIX 25.3   S-PLUS CODE TO PROCESS SIMULATED NONMEM DATA SET

```
#### Import, process, and export simulated data #############
# The simulated data need to be processed to remove observations following the . . .
# . . . first simulated remedication time, if any (i.e. all observartions following . . .
# . . . an observation in which QUIT=1 are removed from the data set)

importDir <- "../"
fileName <- paste(importDir, "analgesicSim.tab", sep="")
simtab <- read.table(fileName, header=T, skip=1)
#
# Determine number of simulated trials, and sample size
nSim <- sum(!duplicated(simtab$ISIM))
nSubj <- sum(!duplicated(simtab$ID))


# Add var to hold logical vector . . . indicating obs to be removed
simtab$REMOVE <- rep(F, nrow(simtab))
```

```
##### Mark observations to be removed
for(iSim in 1:nSim){
  for(iSubj in 1:nSubj){
    TF.subj <- simtab$ISIM==iSim & simtab$ID==iSubj
    TF.quit <- simtab$QUIT[TF.subj]==1
    if(any(TF.quit)){
      Time.v <- simtab$TIME[TF.subj]
      qTime <- min(Time.v[TF.quit])
      simtab$REMOVE[TF.subj] <- simtab$TIME[TF.subj] > qTime
    } # end-if
  } # end-for nSubj
} # end-for iSim


# Remove marked observations
simtab <- simtab[!simtab$REMOVE,]


# Remove records for which the observation is missing (MDV==1 for TIME==0)
TF.v <- simtab$MDV==1
simtab$PRLS[TF.v] <- rep(NA,sum(TF.v))


keepCols <- c("ISIM", "ID", "TIME", "MDV", "PRLS", "QUIT", "PTIM", "DOSE",
              "CL", "VC", "Q", "VP", "KA", "ALAG", "FBIO")


exportDir <- "../"
fileName <- paste(exportDir, "analgesicSim.csv", sep="")


z.export.csv(simtab[, keepCols], fileName, MissingVal=".",AddHash=T)
```

APPENDIX 25.4   S-PLUS CODE CREATED BARPLOTS OF PAIN RELIEF SCORES VERSUS TIME BY DOSE

```
###################################################################
##### Plot Number of Subjects by Time and Pain Relief Score (conditioned on Dose)


##### summarize number of subjects with a given level of pain relief (by dose and time)
n0 <- tapply(simtab$PRLS==0, list(simtab$DOSE, simtab$TIME), sum, na.rm=T)
n1 <- tapply(simtab$PRLS==1, list(simtab$DOSE, simtab$TIME), sum, na.rm=T)
n2 <- tapply(simtab$PRLS==2, list(simtab$DOSE, simtab$TIME), sum, na.rm=T)
n3 <- tapply(simtab$PRLS==3, list(simtab$DOSE, simtab$TIME), sum, na.rm=T)
n4 <- tapply(simtab$PRLS==4, list(simtab$DOSE, simtab$TIME), sum, na.rm=T)
```

```
##### Combine summaries of pain relief scores into a single dataframe
n.all <- rbind(n0, n1, n2, n3, n4)
#
nPRLS <- as.data.frame(n.all)
names(nPRLS) <- dimnames(n1)[[2]]
#
nPRLS$DOSE <- rep(as.numeric(dimnames(n1)[[1]]), 5)
nPRLS$DOSE <- nPRLS$DOSE/1000 # Convert Dose to mg

#### Create stacked barplot of pain relief scores
graphsheet(color.style="black and white")
par(mfrow=c(2,2))
par(cex=1)
#
# Get number of doses, and col that specifies DOSE
Dose.v <- unique(nPRLS$DOSE)
rmCol <- match("DOSE", names(nPRLS))
#
for (iDose in Dose.v){
   TF.v <- nPRLS$DOSE==iDose
  xvals <- barplot(as.matrix(nPRLS[TF.v,-c(1,rmCol)]),ylim=c(0,200),
                 col=seq(5,1), density=500)
  text(xvals, -10, names(nPRLS)[-c(1,rmCol)])
  mtext("Number of Subjects", side=2, line=2.5, cex=1)
  mtext("Time [hr]", side=1, line=1)
  mtext(paste("Dose =", iDose, "mg"))
}
#
key(10,200,
  rectangles=list(size=5, col=seq(5,1), density=500, angle=20),
  text=c("0: none", "1: a little", "2: medium", "3: a lot", "4: complete"),
  title="Pain Relief", cex.title=1.1,
  border=F
)

export.graph("./plots/PRLS.barplot.wmf", ExportType="WMF")
```

APPENDIX 25.5   NONMEM CONTROL FILE TO ESTIMATE PAIN AND REMEDICATION MODEL PARAMETERS

```
$PROB Analgesic Pain Model for nonrandomly censored data


$INPUT ISIM ID=L1 TIME MDV PRLF=DV QUIT PTIM DOSE CL VC Q VP KA ALAG FBIO


; ISIM ID TIME MDV AMT PRLF QUIT PTIM DOSE
; ID = subject ID number
; TIME = time of dose or observation
; MDV = event ID (0=obs, 1=dose)
; AMT = dose amount (MDV=1) or "." (MDV=0)
; PRLF = Ordinal pain relief score (0=no relief thru 4=full relief)
; TQT = Time to remediation
; QUIT = Indicator of remediation (0=stay in study; 1=quit study)
; PTIM = Time of previous pain observation
; DOSE = nominal dose amount


$DATA ../analgesicSim.csv IGNORE=#


$THETA
; Pain Model Parameters
0.4 ; KE0 [1/h]
-2 ; BT1
-2 ; BT2
-2 ; BT3
-2 ; BT4
100 ; EC50 [ng/mL]
20 ; EMAX
0 FIX ; ALPH
2 ; GAMM
5 ; AA
;
; Remediation Model Parameters
0.2 ; LAM0
0.002 ; LAM1
0.002 ; LAM2
0.001 FIX ; LAM3
0.0001 FIX ; LAM4


$OMEGA
; Pain and Remediation Model Parameters
1 ; Pain
0.0 FIX ; Remediation
```

```
$ABBREVIATED DERIV2=NOCOMMON

$PRED
; Specify Pain Model Parameters
KE0 = THETA(1)
BT1 = THETA(2)
BT2 = THETA(3)
BT3 = THETA(4)
BT4 = THETA(5)
EC50 = THETA(6)
EMAX = THETA(7)
ALPH = THETA(8)
GAMM = THETA(9)
AA = THETA(10)
ZPAN = ETA(1)
;
; Specify Remedication Model Parameters
LAM0 = THETA(11)
LAM1 = THETA(12)
LAM2 = THETA(13)
LAM3 = THETA(14)
LAM4 = THETA(15)
ZRMD = ETA(2)

; Calculate concentration in effect compartment
  K20 = CL/VC
  K23 = Q/VC
  K32 = Q/VP
  BET1 = K23+K32+K20
  BET2 = SQRT(BET1**2 - 4*K32*K20)
  BETA = 0.5*(BET1 - BET2)
  ALFA = K32*K20/BETA

  BSL = KE0*KA*DOSE*FBIO/VC
  M1 = ALFA - KA
  M2 = -M1
  Q1 = BETA - KA
  Q2 = -Q1
  R1 = KE0 - KA
  R2 = -R1
  S1 = BETA - ALFA
  S2 = -S1
  T1 = KE0 - ALFA
  T2 = -T1
  U1 = KE0 - BETA
  U2 = -U1
  Z1 = K32 - KA
  Z2 = K32 - ALFA
```

```
  Z3 = K32 - BETA
  Z4 = K32 - KE0


  TIM2 = TIME-ALAG
  E1 = EXP(-KA*TIM2)
  E2 = EXP(-ALFA*TIM2)
  E3 = EXP(-BETA*TIM2)
  E4 = EXP(-KE0*TIM2)


  CE1 = Z1*E1/(M1*Q1*R1)
  CE2 = Z2*E2/(M2*S1*T1)
  CE3 = Z3*E3/(Q2*S2*U1)
  CE4 = Z4*E4/(R2*T2*U2)
IF (TIME .LE. ALAG) THEN
  CE = 0.0
ELSE
  CE = BSL*(CE1 + CE2 + CE3 + CE4)
ENDIF


; Specify placebo effect for each cumulative probability
PEFF = EXP(-ALPH*TIME) - EXP(-GAMM*TIME)
PEFF1 = BT1 + AA*PEFF
PEFF2 = PEFF1 + BT2
PEFF3 = PEFF2 + BT3
PEFF4 = PEFF3 + BT4


; Specify drug effect
DEFF = EMAX * CE/(EC50 + CE)


; Logits for cummulative probabilities
  LGT1 = PEFF1 + DEFF + ETA(1)
  LGT2 = PEFF2 + DEFF + ETA(1)
  LGT3 = PEFF3 + DEFF + ETA(1)
  LGT4 = PEFF4 + DEFF + ETA(1)


; Exponentiate logit
C1 = EXP(LGT1)
C2 = EXP(LGT2)
C3 = EXP(LGT3)
C4 = EXP(LGT4)


; Calculate cumulative probability of response j
  P1 = C1/(1+C1) ; P(Y<=1|X)
  P2 = C2/(1+C2) ; P(Y<=2|X)
  P3 = C3/(1+C3) ; P(Y<=3|X)
  P4 = C4/(1+C4) ; P(Y==4|X)
```

```
; Likelihood (Yj), by pain relief (j = 0 to 4)
 Y0 = 1 - P1
 Y1 = P1 - P2
 Y2 = P2 - P3
 Y3 = P3 - P4
 Y4 = P4


 IND0=0
 IND1=0
 IND2=0
 IND3=0
 IND4=0

IF (PRLF .EQ. 0) IND0=1
IF (PRLF .EQ. 1) IND1=1
IF (PRLF .EQ. 2) IND2=1
IF (PRLF .EQ. 3) IND3=1
IF (PRLF .EQ. 4) IND4=1


; Calculate likelihood of pain score
YP = Y0*IND0 + Y1*IND1 + Y2*IND2 + Y3*IND3 + Y4*IND4


; Calculate likelihood of remedication
LAMM = LAM0
IF(PRLF .EQ. 1) LAMM = LAM1
IF(PRLF .EQ. 2) LAMM = LAM2
IF(PRLF .EQ. 3) LAMM = LAM3
IF(PRLF .EQ. 4) LAMM = LAM4
LAMM = LAMM + ETA(2)


; Prob that subject has not remedicated upto time=TIME
 YR0 = EXP(-LAMM*TIME)
;
; Prob that subject will remedicate at time=TIME, and that subject has
; not remedicated at time=PTIM
 ETIM = TIME - PTIM
 IF(MDV .EQ. 0) THEN
  YR10 = 1 - EXP(-LAMM*ETIM)
 ELSE
  YR10 = 1
 ENDIF
  YR11 = EXP(-LAMM*PTIM)
  YR1 = YR10*YR11

;
; Prob for remedication model
YR = YR0*(1-QUIT) + YR1*QUIT
```

```
; Overall likelihood
IF (MDV .EQ. 0) THEN
  Y = YP*YR
ELSE
  Y = 0
ENDIF


$ESTIMATION SIG=3 MAX=9999 PRINT=1 METHOD=COND LAPLACE LIKE NOABORT

$COV PRINT=ER

$TABLE NOPRINT ONEHEADER FILE=analgesicEst3.tab
  ISIM ID TIME PTIM MDV DOSE QUIT
  P1 P2 P3 P4 Y0 Y1 Y2 Y3 Y4 YR0 YR1 YP YR Y
```

APPENDIX 25.6   S-PLUS CODE CREATED PLOTS OF PROBABILITY/PROPORTION PAIN RELIEF VERSUS TIME BY PAIN RELIEF AND DOSE

```
####################################################################
#### Import PK simulated PK data #############

importDir <- "../"
fileName <- paste(importDir, "analgesicEst.tab", sep="")
esttab <- read.table(fileName, header=T, skip=1)
#
#### Calculate Proportion of subjects with a given level (or
#...greater) Pain Relief...by Dose and Time
prop1<- tapply(esttab$PRLF>=1,list(esttab$TIME,esttab$DOSE),sum,na.rm=T)/
        tapply(!is.na(esttab$PRLF),list(esttab$TIME,esttab$DOSE),sum,na.rm=T)
prop2<- tapply(esttab$PRLF>=2,list(esttab$TIME,esttab$DOSE),sum,na.rm=T)/
        tapply(!is.na(esttab$PRLF),list(esttab$TIME,esttab$DOSE),sum,na.rm=T)
prop3<- tapply(esttab$PRLF>=3,list(esttab$TIME,esttab$DOSE),sum,na.rm=T)/
        tapply(!is.na(esttab$PRLF),list(esttab$TIME,esttab$DOSE),sum,na.rm=T)
prop4<- tapply(esttab$PRLF>=4,list(esttab$TIME,esttab$DOSE),sum,na.rm=T)/
        tapply(!is.na(esttab$PRLF),list(esttab$TIME,esttab$DOSE),sum,na.rm=T)
#
tmp <- data.frame(rbind(prop1, prop2, prop3, prop4))
dose.c <- sort(unique(esttab$DOSE))
names(tmp) <- as.character(dose.c)
#
prop.c <- c("prop1", "prop2", "prop3", "prop4")
#
tmp$PRLF <- rep(prop.c, each=nrow(prop1))
tmp$TIME <- rep(as.numeric(dimnames(prop1)[[1]]), length(prlf.c))
```

```
#
est.prop <- data.frame(make.groups("0"=tmp[,1], "0.5"=tmp[,2], "1"=tmp[,3]))
names(est.prop) <- c("CPROB", "DOSE")
est.prop$DOSE.F <- paste(as.character(est.prop$DOSE), "mg")
est.prop$TIME <- rep(tmp$TIME, length(dose.c))
est.prop$PRLF <- rep(tmp$PRLF, length(dose.c))


#### Calculate Probability of a given level (or greater) Pain Relief by Dose and Time
P1 <- tapply(esttab$P1, list(esttab$TIME, esttab$DOSE), mean, na.rm=T)
P2 <- tapply(esttab$P2, list(esttab$TIME, esttab$DOSE), mean, na.rm=T)
P3 <- tapply(esttab$P3, list(esttab$TIME, esttab$DOSE), mean, na.rm=T)
P4 <- tapply(esttab$P4, list(esttab$TIME, esttab$DOSE), mean, na.rm=T)


tmp <- data.frame(rbind(P1, P2, P3, P4))
dose.c <- sort(unique(esttab$DOSE))
names(tmp) <- as.character(dose.c)


prlf.c <- c("P1", "P2", "P3", "P4")


tmp$PRLF <- rep(prlf.c, each=nrow(P1))
tmp$TIME <- rep(as.numeric(dimnames(P1)[[1]]), length(prlf.c))


est.cProb <- data.frame(make.groups("0"=tmp[,1], "0.5"=tmp[,2], "1"=tmp[,3]))
names(est.cProb) <- c("CPROB", "DOSE")
est.cProb$DOSE.F <- paste(as.character(est.cProb$DOSE), "mg")
est.cProb$TIME <- rep(tmp$TIME, length(dose.c))
est.cProb$PRLF <- rep(tmp$PRLF, length(dose.c))


est.prop.cProb <- rbind(est.prop, est.cProb)

############### Plot Probability/Proportion PRLF vs Time ##################

trellis.device(color=F)

superpose.line.lst <- trellis.par.get("superpose.line")
superpose.line.lst$col <- rep(3:6, 2)
superpose.line.lst$col <- rep(1, 8)
superpose.line.lst$lty <- rep(1:4, 2)
superpose.line.lst$lwd <- rep(3, 8)
trellis.par.set("superpose.line", superpose.line.lst)
superpose.symbol.lst <- trellis.par.get("superpose.symbol")
superpose.symbol.lst$cex <- rep(1, 8)
superpose.symbol.lst$col <- rep(3:6, 2)
superpose.symbol.lst$col <- rep(1, 8)
superpose.symbol.lst$font <- rep(1, 8)
superpose.symbol.lst$pch <- rep(c("1", "2", "3", "4"), 2)
trellis.par.set("superpose.symbol", superpose.symbol.lst)
```

```
tmp.plt <- xyplot(CPROB ~ TIME|DOSE.F, data=est.prop.cProb, groups=PRLF,
  type=rep(c("l","p"), each=4), as.table=T, layout=c(3,1),
  scales=list(cex=1),
  ylab=list("Probability/Proportion Pain Relief >= m", cex=1.2),
  xlab=list("Time [hr]", cex=1.2),
  par.strip.text=list(cex=0.8),
  panel=panel.superpose)

key.lst <- list(text=list(c("m=1", "m=2", "m=3", "m=4"), cex=1, adj=1),
  lines=list(type="o",
    col=superpose.symbol.lst$col[1:4],
    lty=superpose.line.lst$lty[1:4],
    lwd=superpose.line.lst$lwd[1:4],
    pch=superpose.symbol.lst$pch[1:4],
    cex=superpose.symbol.lst$cex[1:4]),
  # x=0.5, y=0.5, corner=c(0,0.5),
  space="top", columns=4, between.columns=2,
  border=1, transparent=F,
  title="Pain Relief", cex.title=1,
  border=1,
  between=1)

# update(tmp.plt, key=key.lst)
update(tmp.plt)
export.graph("./plots/ObsPred.PRLF.TIME.wmf", ExportType="WMF")
```

Transition Models in Pharmacodynamics

ENE I. ETTE

26.1 INTRODUCTION

In diverse fields such as computer science, engineering, mathematics, genetics, agriculture, economics, education, biology, and medical science, random events have been characterized using Markov models (1–3). A Markov model is a stochastic model. Human disease processes, such as diabetic retinopathy (4), systemic lupus erythematosus (5), renal disease (6), papilloma virus and human immunodeficiency virus (7), analgesia (8), and transplantation (9), have been characterized with Markov models. In the above examples Markov models have been used to describe disease as a series of probable transitions between health states. In Chapter 6 of this text the use of a Markov model to characterize patient compliance with prescribed drug therapy is described; it has also been used to characterize sleep patterns (10, 11). The method has considerable appeal for use in pharmacometrics since it offers a method to evaluate patient compliance with prescribed medication regimen, multiple health states simultaneously, and transitions between different sleep stages.

In the subsequent sections an overview of Markov models is provided, followed by a discussion of the Markovian assumption, the discrete time Markov chain, a mixed effects Markov model, and a hybrid mixed effects Markov and proportional odds model suited for data sets that exhibit the characteristics that can be described with such models.

26.2 OVERVIEW OF MARKOV MODELS

A series of probable transitions between states can be described with Markov modeling. The natural course of a disease, for example, can be viewed for an individual subject as a sequence of certain states of health (12). A Markovian stochastic process is memoryless. To predict what the future state will be, knowledge of the current state is sufficient and is independent of where the process has been in the past. This is termed the "strong" Markov property (13).

The characteristics of the state space being measured can be used to classify the Markov process. For most purposes, a discrete or finite space is assumed and this implies that there are a finite number of states that will be reached by the process (14). A continuous or infinite process is also possible. Time intervals of observation of a process can be used to classify a Markov process. Processes can be observed at discrete or restricted intervals, or continuously (15).

Markov chain is the term used to describe a process observed at discrete intervals. However, some investigators prefer to describe Markov chains as a special case of a continuous-time Markov process. That is, the process is only observed at discrete intervals, but in reality it is a continuous-time Markov process (16). Therefore, the Markov process can be used to collectively describe all processes and chains.

Time homogeneity is another important distinction of Markov processes. The process is time independent or time homogeneous when the transition probabilities are constant regardless of the time of observation (12), and the distribution of the number of transitions into a state follows a homogeneous or stationary Poisson process. The Poisson distribution is defined as $P\{N(t) = k\} = (\lambda t^k e^{-\lambda t})/k!$, where λ is the average number of transitions per period t (or the rate of arrivals) over k cycles (17). An exponential distribution defined by the same parameter λ is used to characterize the time between transitions in a homogeneous Poisson process (18).

Short-term medical problems in people are best described with time-homogeneous Markov chains. Chronic disease in people (tens of years) is better described with time-nonhomogeneous models since other factors such as age influence the transition probabilities, therefore causing them to be time dependent (12). A generalization of Markov models can be applied to observations made at irregular time intervals, at irregular intervals with the exact time of transition during that interval unknown, or at regular time intervals (16, 19).

26.3 DISCRETE TIME MARKOV CHAIN

Consider the time-homogeneous model for a disease as exemplified in Figure 26.1, where the transition probabilities are constant over time. Probabilities for the transitions are contained in the transition probability matrix $P(t)$. The probability matrix could simply be written as

$$P = \begin{bmatrix} & H & 2 & 3 \\ H & p_{HH} & p_{H2} & p_{H3} \\ 2 & p_{2H} & p_{22} & p_{23} \\ 3 & p_{3H} & p_{32} & p_{33} \end{bmatrix}$$

since the probabilities for the time-homogeneous model are constant. The rows are representative of the current health state and the columns are representative of the future state. The probabilities are described as p_{ij}, where the probability of moving from state i to state j for any given cycle is p.

When a random variable (potentially) changes states at discrete time points (e.g., every 3 minutes), and the states come from a set of discrete (often, also finite) possible states, a discrete-time Markov chain is used to describe the process.

FIGURE 26.1 Schematic of an example of health/disease stages. A total of six transitions are possible.

Discrete-time Markov chains are discrete-time stochastic processes with a discrete state space. Let the state of the random variable at time t be represented by y_t; then the stochastic process can be represented by $(y_1, y_2, y_3, \ldots)$.

26.3.1 First-Order Markov Chain

With first-order Markov chains, considering all t, the conditional distribution of y_{t+1} given $(y_0, y_1, y_2, \ldots, y_t)$ is identical to the distribution of y_{t+1} given only y_t. That is, we only need to consider the current state in order to predict the state at the next time point. The predictability of the next state is not influenced by any states prior to the current state—the Markov property.

The probability of transition from the ith state at time $t - 1$ to the jth state at time t one step away is given by the conditional probability $P(Y_t = j \mid Y_{t-1} = i) = \pi_{j|i}(t)$. This is called a one-step transition probability. At time t, the process must take on one of the possible states, even if it remains at its current state; therefore, the sum of these probabilities over j is 1.

The joint distribution for a first-order Markov chain depends only on the one-step transition probabilities and on the marginal distribution for the initial state of the process. This is because of the Markov property. A first-order Markov chain can be fit to a sample of realizations from the chain by fitting the log-linear (or a nonlinear mixed effects) model to $[Y_0, Y_1, Y_2, \ldots, Y_{T-1}Y_T]$ for T realizations because association is only present between pairs of adjacent, or consecutive, states. This model states that the odds ratios describing the association between Y_0 and Y_1 are the same at any combination of states at the time points $2, \ldots, T$, for instance.

Consider a disease monitored over a period of 4 months, with evaluations done on a monthly basis. At each evaluation each subject is characterized by the presence (i.e., 1) or absence (i.e., 0) of a symptom. Let the binary response be denoted by Y_t at month t, $t = 1, 2, 3, 4$. As an example, the data on the presence or absence of symptoms over the 4 month period of observation could be summarized as follows:

Y_1	Y_2	Y_3	Y_4	n
0	1	1	1	10
1	1	1	1	15
1	0	1	0	30

where *n* represents the number of subjects who had the same type of symptom score (i.e., 1 or 0) across the 4 months of observation. This type of data could be read into S-Plus® (Insightful, Seattle, WA) and analyzed using the *loglm* model in the MASS library as follows:

```
n.dat<-data.frame(expand.grid(y4=c(1,0), y3=c(1,0), y2=c(1,0),
y1=c(1,0), n=c(10,15,30))
```

A first-order Markov chain (MC) would assume an association between the responses at the first and second months. The prediction of the response for the third month would be independent of the response in the first month, given the response in the second month. Given the response in the third month, the prediction for the fourth month would be independent of the response in the second month. Thus, once the response for a previous month is known, other months are not needed to predict the response for a current month. Data of the type described above could be fitted in S-Plus as follows using the *loglm* function in the S-Plus library MASS:

```
(loglm1.fit<-loglm(n~y1*y2 + y3*y4, data = n.dat,param=TRUE,fit=TRUE))
```

26.3.2 Second-Order Markov Chain

For a second-order Markov model, one would assume that the prediction of response for the fourth month is independent of the response obtained in the first month, given the responses in the second and third months. Such a model could similarly be fitted in S-Plus as

```
loglm2.fit<-loglm(n~y1*y2 *y3+ y2*y3*y4, data = n.dat, param=TRUE,
fit=TRUE))
```

26.4 THE MARKOVIAN ASSUMPTION

The Markovian assumption is easily met in most cases, although it initially appears restrictive. When situations possibly violate the Markovian assumption, adding states to the model may be useful (1, 20, 21). Passage to the state of death in cancer, for example, may occur at a different rate following first remission of the disease than second remission. Using the strategy of adding states, entry into the first of these states forces movement into the next state and there is no backward movement. These states are referred to as tunnel states because they can only be visited in a fixed sequence (20). Adding states increases the complexity of the model and reduces the density of the data for estimation of transition probabilities, but they are helpful in avoiding the violation of the Markovian assumption. There may be limited gain from adding states in the face of loss of precision of estimation of the transition probabilities at some stage in model building.

A discussion of the continuous-time model, the time-nonhomogenous model, and the semi-Markov chain is beyond the scope of this chapter (e.g., see Norris (13),

Stroock (16), Miller and Homan (22), and Yang and Hursch (23) for a discussion on these topics).

26.5 MIXED EFFECTS TRANSITION MODELS

Karlsson et al. (11) used a first-order Markov model to analyze and simulate hypnograms. The purpose of the model was to estimate the probability of moving from one sleep stage to another, and to distinguish the pattern of sleep stage transitions in primary insomniacs (15 women and 6 men) given temazepam (a benzodiazepine) from those of primary insomniacs given a placebo. It was also employed to determine the covariates (nighttime, sleep stage time, and drug exposure) that influence these probabilities. Subjects in the study were determined to be either awake, in rapid eye movement (REM) sleep, or in sleep stages 1, 2, 3, or 4 at 30 second intervals. The Markov model gave the probability of a subject being in a particular stage (e.g., REM sleep) given the fact he/she was in a different sleep stage (e.g., sleep stage 1) in the preceding 30 second interval: that is, $P(REM|stage\ 1)$.

Figure. 26.2 shows six distinct sleep stages with a possible total of 30 transitions from one stage to another. In practice, certain transitions hardly ever occur and not every transition is equally likely; hence, some connections between stages are not shown in Figure 26.2. Karlsson et al. (11) analyzed their data by setting up separate submodels for each transition—the reason being that there were no shared

FIGURE 26.2 Schematic structure of whole night sleep (not drawn to scale). Arrows indicate the directions of possible transitions. A total of 30 (i.e., 6×5) transitions from one sleep stage to another are possible. Some transitions are rare and are therefore not shown in the figure.

parameters between, for example, $P(awake|asleep)$ and $P(asleep|awake)$. Analyzing the data for all subjects simultaneously would introduce correlations that would have to be accounted for.

By assuming a first-order Markov model, Karlsson et al. (11) were able to estimate the probability of, say, REM sleep given sleep stage 1, independently of all other transition probabilities. For simultaneous estimation of transition probabilities for all subjects and to determine the variations in $P(REM|stage\ 1)$ depending on nighttime, stage time, and drug exposure, a mixed effects model was implemented in NONMEM. The parameters of the model were estimated using the conditional likelihood option of the Laplacian method (24). The analysis required the estimation of two parameters—the typical individual value for $P(REM|stage\ 1)$ and the individual distribution of $P(REM|stage\ 1)$. Assuming that the dependent variable (DV) is 1 for REM sleep and 0 otherwise, the critical code of their model was

$$Logit = \log\left[P(REM|stage\ 1)_{\text{typical}}\big/\left(1 - P(REM|stage\ 1)_{\text{typical}}\right)\right] + \eta \qquad (26.1)$$
$$P(REM|stage\ 1)i = \exp(logit)/(1 + \exp(logit))$$

$$Y = P(REM|stage\ 1)i * DV + [1 - P(REM|stage\ 1)i] * (1 - DV) \qquad (26.2)$$

By changing the two parameters of the model, the likelihood of Y was maximized. The model was set up as shown above to allow the estimation $P(REM|stage\ 1)$ instead of its logit transform. The interindividual variability, η, was assumed to be symmetrically distributed with zero mean and a variance ω^2. In modeling the data, the authors had to account for high correlation between the η values.

In characterizing hypnograms using the nonlinear mixed effects modeling approach, it is important to test for correlations between η values of one transition model and those from another model using individual estimates of η values. Correlations detected should be accounted for in the model. Correlations with correlation coefficient (r) ≥ 0.75 are termed high correlations and correlations with r values between 0.5 and 0.75 are moderate correlations (25). Not accounting for such correlations may yield parameter estimates with poor precision.

A detailed description of the nonlinear mixed effects Markov model used to characterize hypnograms following the administration of temazepam can be found in the work of Karlsson et al. (11).

26.6 HYBRID MARKOV MIXED EFFECTS AND PROPORTIONAL ODDS MODEL

Ordered categorical pharmacodynamic data analysis with mixed effects modeling has mostly been performed with the proportional odds model. The model can be used when the data is a categorization of a continuous scale (26) or of a ranking scale (27–33). As discussed in Section 26.1, Markov mixed effects models have been used for characterization of compliance to prescribed medication (19) and for characterizing drug effects on transition between different sleep stages—nonordered

categorical data (11) (see also Section 26.5). In the latter the dependence between the different observations could be properly handled. Zingmark et al. (26) characterized a central nervous system (CNS) adverse effect that occurred suddenly for a new molecular entity under clinical development. This adverse effect occurred suddenly and disappeared and was described as a specific CNS feeling or sensation. The occurrence and severity (none, mild, moderate, or severe) of this specific adverse effect was self-reported by the subjects. A study was performed in which the drug was infused at different rates to mimic absorption profiles of different formulations in an attempt to elucidate the concentration–safety (adverse effect) relationship. The study was conducted in sessions—1 (open session) and 2–4 (double blinded). What made the data collected in that study different from the general ordered categorical side effect data reported in the literature is the continuous monitoring of both times and type of severity of the change of stage of the adverse effect. The exact time, severity, and type of change in severity of the adverse effect were recorded. The adverse event was classified as 0 (no adverse effect), 1 (mild adverse effect), and 2 (moderate or severe adverse effect). All subjects started their treatment without the adverse effect: that is, the baseline observation was 0. Two types of events had the possibility of occurring at any one time. Thus, the data consisted of times and events: six types of events occurred with all possible transitions between three stages. The transitions realized were from 0 to >1, from 0 to >2, from 1 to >2, from 2 to >1, from 2 to >0, and from 1 to >0.

Given the graded nature of the adverse effect score and the occurrence of transitions between scores, Zingnmark et al. (26) represented the data as observations of severity stage on an equal-spaced time grid, with observations frequent enough to adequately represent the data. A 3 minute interval between observations was selected on an ad hoc basis to characterize the transitions. This was felt to be sufficiently frequent enough so that in no interval did more than one transition between severity stages occur. The time interval was considered adequate for the application of the model.

A standard proportional odds model for ordered categorical data described in Chapter 25 cannot be used to analyze frequently measured categorical type data that have a pronounced correlation between neighboring observations. The model discussed in Chapter 25 assumes that observations are independent. Transition models including Markov elements discussed in the previous sections have been used for such situations (11). The data set used by Zingmark et al. (26) was not sufficiently information rich to allow appropriate estimation of all resulting parameters that are characteristic of transition models—one model for each transition. Thus, they used a *hybrid* of a proportional odds model and a transition model. That is, they accounted for the ordered nature of the data in the parameterization of the model and also accounted for transitions between stages (i.e., estimation of the probabilities of having a certain adverse effect score conditioned on a preceding observation). The model can also be viewed as a proportional odds model with a first-order Markov chain. The model was used to obtain estimates of the cumulative probabilities of the adverse effect scores, given a preceding observation. Letting an observed adverse effect score be denoted by S, and given a preceding observation ($pre = 0, 1,$ or 2), the logits (lx) of the probabilities that $S \geq 1$ or $S = 2$ were described by the following:

$$[l_S \geq 1 | pre = 0] = b_1 + D$$

$$[l_S = 2 | pre = 0] = b_1 + b_2 + D$$

$$[l_S \geq 1 | pre = 1] = b_3 + D$$

$$[l_S = 2 | pre = 1] = b_3 + b_4 + D \qquad (26.3)$$

$$[l_S \geq 1 | pre = 2] = b_5 + D$$

$$[l_S = 2 | pre = 2] = b_5 + b_6 + D$$

where baseline fixed effects parameters of the model are denoted by b_x and D is the drug effect. Different formulations of drug effect can be used. Equation (26.4) provides the probabilities corresponding to the logits in Eq. (26.3):

$$PC_x = e^{lx} / 1 + e^{lx} \qquad (26.4)$$

Thus, actual probability, p_x, of observing a particular score is

$$p_{S=0} = 1 - PC_{S \geq 1}, \quad p_{S=1} = PC_{S \geq 1} - PC_{S=2}, \quad p_{S=2} = PC_{S=2} \qquad (26.5)$$

The hybrid model proposed by Zingmark et al. (26) is a straightforward way of incorporating Markov elements in an analysis of ordered categorical data. An inappropriate model—a bad descriptive model or a model with a bad predictive performance (see Ette et al. (34); Chapter 8 of this text)—would result if the correlated nature of the data is ignored and a proportional odds model is used to characterize the concentration–adverse effect relationship. Readers are referred to the article by Zingmark et al. (26) for a detailed description of the hybrid model. They also provide a NONMEM data set and control file for the implementation of the model.

26.7   SUMMARY

Markov models are used to describe disease as a series of probable transitions between health states. The methodology has considerable appeal for use in pharmacometrics since it offers a method to evaluate patient compliance with prescribed medication regimen, multiple health states simultaneously, and transitions between different sleep stages. An overview of the Markov model is provided together with the Markovian assumption. The most commonly used form of the Markov model, the discrete-time Markov model, is described as well as its application in the mixed effects modeling setting. The chapter concludes with a discussion of a hybrid Markov mixed effects and proportional odds model used to characterize an adverse effect that lends itself to this combination modeling approach.

REFERENCES

1. A. Hillis, M. Maguire, B. S. Hawkins, and M. M. Newhouse, The Markov process as a general method for nonparametric analysis of right-censored medical data. *J Chron Disease* **39**:595–604 (1986).

2. S. Jain, Markov chain model and its applications. *Comp Biomed Res* **19**:374–378 (1986).

3. W. J. Stewart, *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, 1994.

4. G. Marshall and R. H. Jones, Multi-state models and diabetic retinopathy. *Stat Med* **14**:1975–1983 (1995).

5. M. D. Silverstein, D. A. Albert, N. M. Hadler, and M. W. Ropes, Prognosis in SLE: comparison of Markov model to life table analysis. *J Clin Epidemiol* **41**:623–633 (1988).

6. D. E. Schaubel, H. I. Morrison, M. Desmeules, D. Parsons, and S. S. A. Fenton, End-stage renal disease projections for Canada to 2005 using Poisson and Markov models. *Int J Epidemiol* **27**:274–281 (1998).

7. J. C. M. Hendriks, G. A. Satten, I. M. Longin, H. A. M. van Druten, P. T. A. Schellekens, R. A. Coutinho, and G. J. P. van Griensven, Use of immunological markers and continuous-time Markov models to estimate progression of HIV infection in homosexual men. *AIDS* **10**:649–656 (1996).

8. K. Meyer, Clinical trial factors in a pain transition state model, in *Proceedings of the 2005 Winter Simulation Conference*, M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines (Eds.). 2005, pp. 2232–2235.

9. N. Keiding, J. P. Klein, and M. M. Horowitz, Multi-state models and outcome prediction in bone marrow transplantation. *Stat Med* **20**:1871–1885 (2001).

10. B. Kemp and H. A. C. Kamphuisen, Simulation of human hypnograms using a Markov chain model. *Sleep* **9**:405–414 (1986).

11. M. O. Karlsson, R. C. Shoemaker, B. Kemp, A. C. Cohen, J. M. A. van Gerven, B. Tuk, C. C. Peck, and M. Danof, A pharmacodynamic Markov mixed-effects model for the effect of temazepam on sleep. *Clin Pharmacol Ther* **68**:175–188 (2000).

12. J. R. Beck and S. G. Pauker, The Markov process in medical prognosis. *Med Decis Making* **3**:419–458 (1983).

13. J. R. Norris, *Markov Chains*. Cambridge University Press, Cambridge, UK, 1997.

14. P. A. Jensen and J. F. Bard, Mathematics of discrete-time Markov chains, in *Operations Research Models and Methods*. Wiley, Hoboken, NJ, 2002, pp. 466–492.

15. O. Kempthorne and L. Folks, Markov chains, in *Probability, Statistics, and Data Analysis*. Iowa University Press, Ames, 1971, pp. 188–197.

16. D. W. Stroock, *Introduction to Markov Process* (Graduate Texts in Mathematics). Springer, New York, 2005.

17. P. A. Jensen and J. F. Bard, Sensitivity analysis, duality and interior point methods, in *Operations Research Models and Methods*. Wiley, Hoboken, NJ, 2002, pp. 111–145.

18. W. Q. Meeker and L. A. Escobar, Models for continuous failure time processes, in *Statistical Methods for Reliability Data*. Wiley, Hoboken, NJ, 1998, pp. 26–45.

19. P. Girard, T. F. Blaschke, H. Kastrissios, and L. B. Sheiner, Markov mixed effect regression model for drug compliance. *Stat Med* **17**:2313–2333 (1998).

20. F. A. Sonnenberg and J. R. Beck, Markov models in medical decision making: a practical guide. *Med Decis Making* **13**:322–338 (1993).

21. A. R. de Kruyk, J. H. P. vander Meulen, L. A. van Hererden, J. A. Bekkers, E. W. Steyerberg, R. Dekker, and J. D. F. Habbema, Use of Markov series and Monte Carlo simulation in predicting replacement valve performances. *J Heart Valve Dis* **7**:4–12 (1998).

22. D. K. Miller and S. M. Homan, Determining transition probabilities: confusion and suggestions. *Med Decis Making* **14**:52–58 (1994).

23. M. C. Yang and C. J. Hursch, The use of semi-Markov model for describing sleep patterns. *Biometrics* **29**:667–676 (1973).

24. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. NONMEM Project Group, University of California, San Francisco, 1992–1998.

25. E. I. Ette, A. W. Kelman, C. A. Howie, and B. Whiting, Interpretation of simulation studies for efficient estimation of population pharmacokinetic parameters. *Ann Pharmacother* **27**:1034–1039 (1993). Erratum in *Ann Pharmacother* **27**:1548 (1993).

26. P.-H. Zingmark, M. Kagedal, and M. O. Karlsson, Modelling a spontaneously reported side effect by use of a Markov mixed-effects model. *J Pharmacokinet Pharmacodyn* **32**:261–281 (2005).

27. L. B. Sheiner, A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther* **56**:309–322 (1994).

28. C. Y. Liu and N. C. Sambol, Pharmacodynamic analysis of analgesic trials using empirical methods. *Pharm Res* **12**:438–445 (1995).

29. J. W. Mandema and D. R. Stanski, Population pharmacodynamic model for ketorolac analgesia. *Clin Pharmacol Ther* **60**:619–635 (1996).

30. L. B. Sheiner, S. L. Beal, and A. Dunne, Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *J Am Stat Assoc* **92**:1235–1255 (1997).

31. J. W. Mandema, Population pharmacokinetics/pharmacodynamics of analgesics: theory and applications, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, M. Danhof, M. Gex-Fabry, U. A. Gundert-Remy, M. O. Karlsson, F. Mentre, P.L. Morselli, F. Rombout, M. Rowland, J. L. Steimer, and S. Vozeh (Eds.). Commission of the European Communities, European Cooperation in the Field of Scientific and Technical Research, Brussels, 1997, pp. 74–82.

32. E. I. Ette, P. Lockwood, R. Miller, and J. W. Mandema, Analysis of analgesic trials, in *Applied Statistics in the Pharmaceutical Industry Using S-Plus*, S. Millard and A. Krause (Eds.). Springer-Verlag, New York, 2001, pp. 237–266.

33. P.-H. Zingmark, M. Ekblom, T. Odergren, T. Ashwood, P. Lyden, M. O. Karlsson, and E. N. Jonsson, Population pharmacokinetics of clopmethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol* **56**:173–183 (2003).

34. E. I. Ette, P. J. Williams, Y. K. Kim, E. V. Cappazelli, The determination of population pharmacokinetic model appropriateness, *J Clin Pharmacol* **43**:2–15 (2003).

Mixed Effects Modeling Analysis of Count Data

CHRISTOPHER J. GODFREY

27.1 INTRODUCTION

The use of pharmacokinetic/pharmacodynamic (PK/PD) modeling for the analysis of continuous variables in clinical studies is well established and widely accepted. This is regardless of whether the effect of drug exposure acts directly or indirectly on some measured PD endpoint. Models characterizing direct PD responses are the most widely used. However, there is a growing body of literature reports on the use of indirect response PD models for continuous endpoints.

Far fewer PK/PD analyses and reports deal with PD measures that are discrete in nature. Variables that represent discrete measures have properties that are distinct from continuous variables. A variable is discrete if the number of values that it can assume is finite or countably infinite. Count data is generally considered a type of discrete variable.

Count data describe the number of times some event of interest occurs in some specified time frame or over some one- or multidimensional spatial distance. Count data arise in drug development programs and clinical studies that seek to characterize the effect of a pharmacologic intervention on the occurrence frequency of undesirable events. Examples of endpoints of interest include seizures in epilepsy (1, 2), angina episodes in acute coronary syndromes, migraine headaches, apneic episodes in the premature neonate (3), asthma attacks, the ipecac model of chemotherapy-induced vomiting (4), premature ventricular contractions (5), urge incontinence episodes (6), and panic attacks. These events are observed on a time scale and are frequently expressed as the number of occurrences in a clinically relevant time frame, for example, apneic episodes per day or panic attacks per month. Examples of spatially expressed count data may include the number of polyps in a specified length of the colon (7, 8) or psoriatic plaques in a defined area of skin. Count data may also arise in examining adverse events documented in a clinical investigation. Note that in each case the endpoint is characterized by a specific and finite event.

While the severity of some events may be quantified on a quasicontinuous scale, such as a visual analog scale, the focus of the present treatment deals with the event rate.

Frequentist-based statistical methodology is available for the analysis of count data, including, for example, Poisson regression. However, a PK/PD modeling approach to the analysis of count data has distinct advantages. The development of parametric PK/PD models permits quantification of exposure–response for count data. Such models may readily facilitate investigation of possible explanatory variables, simultaneous characterization of time and exposure relationships, and simulation of future study and development program outcomes. Additionally, models may permit a mechanistic interpretation of drug impact on the endpoint of interest.

An additional advantage is the opportunity to conduct PK/PD/outcome modeling. In this case the outcome refers to the occurrence rate of the event of interest, and the PD component is a biomarker that is, presumably, predictive of or associated with the event occurrence. An example would be a model for a drug that reduces inflammation, as measured by the inflammatory biomarker C-reactive protein, and thus reduces the incidence of angina episodes in acute coronary syndromes. C-reactive protein is the link between drug exposure and angina frequency. Such PK/PD/outcome models potentially have significant benefit for drug development. Rigorous and qualified models may contribute to the validation of the biomarker, aid in the development of a diagnostic, or form the basis of model-driven simulation.

The aim of this chapter is to equip the pharmacometrician with sufficient theory and application to confidently approach the PK/PD-based analysis of count data and thus derive the maximum return on investment from clinical study data. Section 27.2 provides a motivating example and Section 27.3 presents relevant definitions and theory. Section 27.4 applies the theory to the example and introduces diagnostics methods. Throughout the chapter, the focus is on population approaches using nonlinear mixed effects models. Code segments of NONMEM control files are presented in the appendix. Mixed effects analysis methodology is described in detail in Chapter 4 of this text.

27.2 MOTIVATING EXAMPLE: NEONATAL APNEA

Apnea of prematurity is a common disorder afflicting neonates less than 37 weeks gestation. Apnea is characterized by cessation of respiration for a duration ≥15 seconds with or without accompanying bradycardia (≤80 bpm for 10 s). Pharmacologic intervention typically centers on administration of methylxanthines, either theophylline or caffeine. While a number of manuscripts have reported the population pharmacokinetics of theophylline (9–11), scant information is available on models for theophylline activity in neonatal apnea. An investigation was conducted to determine the population pharmacokinetics and pharmacokinetics/pharmacodynamics of orally and intravenously administered theophylline in premature neonates using a routine clinical care study design: that is no constraints were placed on the dosing or observation schedule for drug concentration. Blood samples for measurement of theophylline were obtained by heel stick or indwelling catheter at a frequency determined according to standard hospital protocol. Infants were

FIGURE 27.1   Distribution of apneic episodes per day. The fraction of all patient days is represented by each level of apneic episodes.

continuously monitored for apnea through a centralized computer tracking respiratory rate, heart rate, and oxygenation. Apneic and bradycardic episodes triggered an audible alarm at the central station and bedside computer monitor. The episode was hand-recorded in a bedside apnea logbook by the caregiver.

The study enrolled 97 infants of gestational ages of 24–33 weeks admitted to the neonatal intensive care unit of Connecticut Children's Medical Center (Hartford, CT) (3). Twenty-eight infants were studied prospectively, 69 retrospectively. The neonates were studied from birth up to a maximum of 18.286 postnatal weeks. Data on daily apneic episodes were obtained on approximately 5000 patient-days for 95 of the neonates. Figure 27.1 shows a distribution of the number of apneic episodes per day across all neonates and postnatal days studied. The mean (SD) number of spells per day was 1.93 (2.91). The range was 0–30 episodes per day. Figure 27.2 depicts the time course of the mean daily episode count with respect to postnatal age. The frequency increases up to approximately 1.5 weeks and declines gradually as the infants mature.

A population PK and PK/PD analysis was performed to develop a model for the time course of theophylline concentrations and for the time course and exposure–response of apneic episodes to treatment with theophylline (3). Results of the population pharmacokinetics of theophylline will not be presented.

27.3   THEORY ON THE ANALYSIS OF COUNT DATA

This section covers introductory theory about the distribution and analysis of count-type data. The focus is on the Poisson and zero-inflated Poisson distributions. The mechanics for implementation in an analysis are also discussed.

FIGURE 27.2   Mean (±SD) apneic episodes per day by postnatal age. Error bars are +1 standard deviation. Postnatal days for which episodes were recorded in only one neonate are excluded.

27.3.1   Poisson Data Distribution Model

Since distributions describing a discrete random variable may be less familiar than those routinely used for describing a continuous random variable, a presentation of basic theory is warranted. Count data, expressed as the number of occurrences during a specified time interval, often can be characterized by a discrete probability distribution known as the Poisson distribution, named after Simeon-Denis Poisson who first published it in 1838. For a Poisson-distributed random variable, Y, with mean λ, the probability of exactly y events, for $y = 0, 1, 2, \ldots$, is given by Eq. (27.1). Representative Poisson distributions are presented for $\lambda = 1$, 3, and 9 in Figure 27.3.

$$P(Y = y; \lambda) = \frac{e^{-\lambda} \times \lambda^y}{y!} \quad \text{for } \lambda > 0 \text{ and } y = 0, 1, 2, \ldots \quad (27.1)$$

The Poisson distribution tends to symmetry as λ increases. For $\lambda > 10$, the Poisson distribution is reasonably well represented by a normal distribution. This has implications for analysis in cases where the mean number of counts is expected to be high, in which case traditional analyses for continuous data may be sufficient.

A key feature of a Poisson-distributed random variable is that it is completely described by one parameter, λ. For the Poisson distribution, the variance is equal to the mean. However, clinical count data often can exhibit overdispersion, where the variance exceeds the mean. In this case, the variance of Y, $\text{Var}(Y)$, equals $\phi\lambda$, where ϕ is the overdispersion parameter. A number of alternative distributions can be used to describe overdispersed data, such as the negative binomial (12).

FIGURE 27.3 Simulated Poisson distribution for means 1, 3, and 9. Histograms depict the distribution of a Poisson distributed random variable with means of 1, 3, and 9.

Mixed effects modeling approaches provide a natural way to account for overdispersion. Instead of assuming all units have a common mean (i.e., a homogeneous Poisson), a distribution of means is assumed (i.e., a heterogeneous Possion). Thall (13) proposed a mixed gamma–Poisson, where the Poisson means follow a gamma distribution. Frame et al. (14) employed a lognormal Poisson in their analysis of partial seizure frequency, as did Gupta et al. (6) in the analysis of urge incontinence. As will be shown, nonlinear mixed effects modeling allows for and accounts for overdispersion.

The distribution presented in Eq. (27.1) is an appropriate distributional description if the data from each study subject is characterized by the same mean and observations are homogeneous across time. However, it is more typically the case that each subject may have observations arising from individual Poisson distributions. Therefore, let $y_{i,j}$ represent the number of events observed in the ith subject during the jth interval. It follows that

$$P(Y_{i,j}; \lambda_i) = \frac{e^{y_{i,j}} \times \lambda_i}{y_{i,j}!}$$

$$\lambda_i = f(x_{i,j}, \theta, \eta_i)$$

(27.2)

where θ is a vector of fixed effects parameters, $x_{i,j}$ is a vector of subject-specific and possibly time-specific parameters, and η is an interindividual random effects parameter characterizing the deviation of the ith subject's estimate from the population mean estimate. η is a normally distributed random variable with mean 0 and variance ω^2. The mean may depend on time, usually arising as a result of the underlying temporal pattern of disease. The functional relationship, $f(.)$, is unspecified here

but may take many forms. Typically, $f(.)$ is an exponential function of an equation linear in the predictors. Thus, under a log-linear canonical link,

$$\xi_{i,j} = \log(\lambda_{i,j}) = \theta x_{i,j} + \eta_i \qquad (27.3)$$

where ξ is the log of the mean, θ is the vector of fixed effects parameters, x is the vector of independent variables, and η is as described previously. This expression has an attractive property of constraining the mean to be positive and nonzero, a requirement for the Poisson. However, interpretation of the model parameters may be less than straightforward under this transformation.

Consideration of the underlying time course is potentially quite important in the presence of disease progression, disease resolution, or remitting–relapsing progression (see Chapter 21 for a discussion of disease progression models). However, sufficient data obtained from control groups (placebo, active, or historical) may be needed to adequately elaborate this element of the model.

27.3.2 Zero-Inflated Poisson Data Distribution Model

Another situation that may be observed with count data in clinical studies is "zero-inflation." This condition is characterized by the occurrence of a higher frequency of time intervals with no events than would be expected under a Poisson distribution. While zero-inflation may be adequately addressed by allowing overdispersion through inclusion of interindividual variability, it may also represent a phenomenon distinct from overdispersion. Under such cases, the zero-inflated Poisson (ZIP) may be a useful distribution to use in analysis. The ZIP can best be understood from a contrived, if not unrealistic, example. If one was going to test the activity of an antiallergy medicine by counting the number of sneezes during 8 hour intervals, one might find that there were more intervals with no sneezes than expected from a Poisson-distributed variable. An underlying explanation, albeit possibly unobserved, could be that subjects might not have been exposed to the offending allergen in all of those intervals. For example, time spent at work might be time away from a pet at home or pollen outdoors. In some intervals, subjects will not sneeze since they are not being exposed to allergen; therefore, they are considered to be in a "perfect" or nonsusceptible state. The first question that one can ask is: What is the probability that a subject is in a perfect state or imperfect state? In a perfect state, the event of interest is not observed. If a subject is likely to be in the imperfect state, one can ask what the probability is of experiencing 1, 2, or any number of events.

A ZIP distribution is a mixing distribution of a Bernoulli and a Poisson distribution, and using it in analysis is like combining a logistic regression with Poisson regression. The derivation, properties, and application of the ZIP have been presented in a number of papers (e.g., 15–17). There are two ways to structure the ZIP in analysis that should give the same result but may differ in interpretation of the parameters. The choice will also influence how the likelihood function is coded in the model file. In the first approach, the logistic component will provide the probability of being in the perfect or imperfect state. If the probability is such that the subject is considered to be in the imperfect state, the full Poisson is used to determine the probability of observing the number of events. To determine

the probability of observing zero events, one must combine the probability of the perfect state, in which only zero events are allowed, and the probability of having zero events given the imperfect state. Thus,

$$P(Y = 0; \phi, \lambda) = \phi + (1 - \phi) e^{-\lambda} \tag{27.4}$$

$$P(Y = y; \phi, \lambda) = (1 - \phi) \frac{e^{-\lambda} \times \lambda^y}{y!} \quad \text{for } y = 1, 2, 3, \ldots \tag{27.5}$$

where ϕ is the probability of the perfect state and contributes the excess zeros. The second term in Eq. (27.4) is the probability of zero arising from the Poisson characterizing the imperfect state. Implicit in this formulation is that zero events can occur even when the imperfect or susceptible state is the most likely. In the second approach, the logistic function is used to characterize the probability of having zero events, whether the subject is in the perfect or imperfect state. A truncated Poisson is used to characterize the probability of nonzero event rates. This approach is also call the Hurdle model (16). In the Hurdle model formulation,

$$P(Y = 0; \phi) = \phi \tag{27.6}$$

$$P(Y = y; \phi, \lambda) = (1 - \phi) \frac{e^{-\lambda} \times \lambda^y}{y!} \left(1 - e^{-\lambda}\right)^{-1} \quad \text{for } y = 1, 2, 3, \ldots \tag{27.7}$$

ϕ is simply the probability of observing zero events, regardless of the state. Equation (27.7) is a truncated Poisson, where the probability of observing zero events is in effect removed.

With either formulation, the state probability ϕ and the Poisson mean λ can each potentially include a random effect for interindividual variability. With zero-inflated Poisson-based analyses, covariates can be evaluated as predictors of the probability of state, the mean number of events, or both. Under typically employed canonical links,

$$\xi_{i,j} = \log(\lambda_{i,j}) = \theta x_{i,j} + \eta_i \tag{27.8}$$

as in the regular Poisson model and

$$\gamma_{i,j} = \log \text{it}(\phi_{i,j}) = \theta x_{i,j} + \eta_i \tag{27.9}$$

where the logit is

$$\log \text{it}(\phi_{i,j}) = \log\left(\frac{\phi_{i,j}}{1 - \phi_{i,j}}\right) \tag{27.10}$$

The two components may share elements of $x_{i,j}$. However, the vectors θ and η are disjoint.

Formulas are available for the computation of the expected value and variance of the ZIP random variable (18). The expected value for a ZIP-distributed random variable Y, $E(Y)$, is

$$E(Y) = (1 - \phi)\lambda \tag{27.11}$$

The variance of Y, $\text{Var}(Y)$, is calculated as follows:

$$\text{Var}(Y) = E(Y) + E(Y)[\lambda - E(Y)] \tag{27.12}$$

These formulas are useful when seeking to diagnose the adequacy of model fit.

27.3.3 Implementation

Utilization of the Poisson and ZIP in population PK/PD modeling requires coding the appropriate distribution into the software selected for analysis. Example code will be given as appropriate for NONMEM implementation; however, the fundamentals are applicable to other software programs.

Typically the $PRED block will be used to code all elements of the model. However, use of PREDPP and the associated ADVAN subroutines is possible, particularly when a combined PK/PD model is desired (e.g., see Ref. 4). For this discussion, a model for exposure–response based on the Poisson distribution is assumed and area under the concentration–time curve (AUC) is the exposure metric. The first part of the model provides for definition of the mean and variance of the Poisson distribution.

```
$PRED
TLLAM=THETA(1)+THETA(2)*AUC ;Typical value of log(lambda)
LLAM = TLLAM +ETA(1)  ;Add intersubject variability
LAM = DEXP(LLAM)  ;Inverse canonical link
```

If LAM is expressed as a linear function, instead of using the log link, care must be taken to appropriately constrain the parameters such that LAM is always positive and nonzero. LLAM can assume more complex forms as needed to adequately characterize the data. If a ZIP is to be used, the following code is added:

```
TLOGIT = THETA(3) ;Typical value of the logit
LOGIT = TLOGIT + ETA(2) ;Add intersubject variability
PHI = DEXP(LOGIT)/(1+DEXP(LOGIT)) ;Inverse canonical link
```

Again, using a linear function to describe PHI can create difficulties unless constraints are used to bound PHI between 0 and 1 at the subject level.

The second part of coding the model provides the likelihood function for the Poisson or ZIP distribution. A number of options need to be considered when creating this part of the control file. One needs to decide whether to write the likelihood function or $-2 \times$ log likelihood function. Either is acceptable but NONMEM must be provided the correct designation (LIKE or -2LL) in the $ESTIMATION block.

Another issue to consider is that the factorial part of the Poisson distribution cannot be coded directly as a factorial. Two options exist for addressing this. The simplest is to include the value of the factorial of the observation in the data set and include the variable in the $INPUT block. Most software that the pharmaco-

metrician might use to create a data set can easily handle calculation of the factorial for inclusion in the data set. Another option is to approximate the factorial using Stirling's formula, which is based on the gamma function. The formula is as follows:

$$y! = (2\pi y)^{1/2}\, y^y \exp(-y) \quad \text{for } y = 1, 2, 3, \ldots \qquad (27.13)$$

and

$$y! = 1 \quad \text{for } y = 0 \qquad (27.14)$$

NONMEM code for Stirling's formula would thus appear as

```
IF(DV.EQ.0) THEN
FACT=1
ELSE
FACT=((6.28312*DV)**(0.5))*(DV**DV)*DEXP(-DV)
ENDIF
```

The log factorial can easily be derived from the above expressions if the −2LL option is used. A word of caution is warranted if Stirling's formula is employed. When the mean number of counts is small and the number of observed counts is small, the Possion probability calculated using Stirling's formula could be off by a few percentage points. For example, when the Poisson mean is 1, the probability of observing 1 event is approximately 0.368. However, under a Poisson calculated with Stirling's formula, the probability is nearly 0.4. While this difference is not large, it may be relevant if one is comparing two treatments and one treatment is more effectively driving the number of events toward zero than the other. The bias introduced by Stirling's formula may result in an optimistic estimation of drug effect.

NONMEM code for the Poisson likelihood appears as follows:

```
POIS=LAM**DV*DEXP(-LAM)/FACT
Y=POIS
```

where FACT is either a data item or Stirling's formula.

The ZIP likelihood can be combined with the Poisson likelihood as follows:

```
STATE=0
IF (DV.EQ.0) STATE=1          ;Set indicator variable
P0=PHI + (1-PHI)*DEXP(-LAM) ;Probability of zero count
PN=(1-PHI)*POIS               ;Probability of count 1,2,3,...
ZIP=P0**STATE * PN**(1-STATE)
Y=ZIP
```

With the basic theory and implementation in NONMEM covered, these methods can be employed in the analysis of the data described in Section 27.2.

27.4 APPLICATION OF POISSON-BASED POPULATION ANALYSIS TO APNEIC EPISODE DATA

This section demonstrates aspects of the application of a nonlinear mixed effects modeling approach to the analysis of count data using the premature neonate apnea data described in Section 27.2. The objective is to draw attention to key features the pharmacometrician should be aware of and provide methods for model diagnostics and general considerations. Selected results presented here are excerpted and adapted from the complete analysis (3). A subset of the analysis data set is provided in the appendix.

27.4.1 Model for the Distribution of the Data

The first step is to implement a Poisson model characterized by a mean with one interindividual variance term (see Model 1 in the appendix). The estimate of the population λ was approximately 1.42 episodes per day with a 74% interindividual coefficient of variability (CV). Although not executed in the original analyses, for comparison purposes, a zero-inflated Poisson model was also evaluated (Model 2 in the appendix). The estimated probability of the nonsusceptible state was 0.342 (71% interindividual CV). The ZIP population λ estimate was 2.33 (55% interindividual CV) episodes per day. Using the formulas given in Eq. (27.11) and (27.12), the population expectation (variance) was 1.53 (2.76) episodes per day. Note the overdispersion as indicated by the variance greater than the expected value. Figure 27.4 shows the predicted probability from the Poisson and the ZIP

FIGURE 27.4   Poisson and zero-inflated Poisson estimated probability versus observed frequency. The solid circles are the estimated probabilities of observing the episode rate from the Poisson distribution. The open squares are the estimated probabilities of observing the episode rate from the zero-inflated Poisson distribution. The histogram is the distribution of the observed data.

superimposed on the observed frequency of episode counts. Both models exhibit deficiencies in characterizing the full nature of the observed data; however, the ZIP model does perform as intended and accounts for the extra zeros. Since the two models use different objective functions, it is inappropriate to compare objective function values for model selection. Despite the potential advantage of the ZIP model, the remainder of this example uses the Poisson model. As will become apparent, the Poisson model can accommodate the zero-inflation and some of the overdispersion present once the apnea time course and covariates have been incorporated into the model.

27.4.2 Time Course Model

Apnea of prematurity generally resolves as the neonate matures and approaches typical term age. Therefore, the PK/PD model must account for this trend with respect to time. Figure 27.2 showed the time course of apneic episode frequency. The number of daily episodes increased after birth with a peak between 1 and 2 weeks on average. A gradual decline was observed thereafter. A number of functional forms were considered to describe this profile, but two of the models evaluated are presented here. The first model, TC1 (Eq. (27.15)), included a zero-order progression rate of episode frequency and a first-order resolution rate of episode frequency. In the second model, TC2 (Eq. (27.16)), the progression and resolution rates were both treated as first-order processes. Note that the use of resolution is meant to imply lessening of disease severity with maturity, not resolution of a specific apneic episode.

$$\text{Model TC1}: \quad \lambda_{i,j} = \left[\frac{k_0}{K} \left(e^{K \times T} - 1 \right) e^{-K \times PNA_{i,j} \times 7} \right] e^{\eta_{1i}} \qquad (27.15)$$

where $\lambda_{i,j}$ is the ith subject-specific mean on the jth postnatal day, k_0 is the zero-order progression rate of apneic episodes, K is the first-order resolution rate of apnea, and T represents the time to maximum spell frequency. $PNA_{i,j}$ is the jth postnatal age, converted from weeks to days, in the ith subject. The η_{1i} is the ith subject-level realization of the intersubject random effects term, η_1, a symmetrically distributed random variable with mean 0.

$$\text{Model TC2}: \quad \lambda_{i,j} = \frac{PRE \times K_{\text{in}}}{K_{\text{in}} - K_{\text{out}}} \left(e^{-K_{\text{out}} \times PNA \times 7_{i,j}} - e^{-K_{\text{in}} \times PNA \times 7_{i,j}} \right) e^{\eta_{1i}} \qquad (27.16)$$

where PRE is a preexponential term, K_{in} is the first-order progression rate of spells, and K_{out} is the first-order resolution rate.

Both models resulted in a substantial improvement in model fit. The reduction in the $-2\times$ log likelihood objective function value from the model without a time course component was 1876 for Model TC1 and 2031 for Model TC2. Model TC2 was selected as the time course model.

27.4.3 Covariate Model

An advantage of the particular functional form selected for the time course is that covariates can be tested on the parameters that characterize the time course. More

specifically, a covariate could be a predictor of how quickly apnea increases in severity, the maximum frequency of the apnea, or the resolution of the condition. Covariates that were considered as possible explanatory factors included gestational age (GA), continuous positive airway pressure (CPAP, 0 = absent, 1 = present), gender (GEN, 0 = male, 1 = female), maternal prenatal steroid administration (MSTR, 0 = not administered, 1 = administered), race/ethnicity (RACE), hyaline membrane disease (HMD, 0 = absent, 1 = present), bronchopulmonary dysplasia (BPD, 0 = absent, 1 = present), concurrent infection (INF), and APGAR score at 1 and 5 minutes. Covariate selection was accomplished by first triaging covariates through a screening procedure then using a forward selection followed by backward elimination process. The elimination process was performed after the exposure–response model was added as described below. The order of covariate inclusion was governed by the significance level calculated using the likelihood ratio test obtained during the screening process. (Chapters 8 and 14 of this text provide a more thorough treatment of covariate selection and model building approaches.) Retention of a covariate during the forward selection process required statistical significance at the 0.05 level. All covariates were entered linearly in the model. The complete covariate model was as follows:

$$
\begin{aligned}
PRE_i &= \theta_1 + \theta_2\,(GA_i - 24) \\
K_{\mathrm{in},i} &= \theta_3 + \theta_4\,(GA_i - 24) \\
K_{\mathrm{out},i} &= [\theta_5 + \theta_6\,(GA_i - 24)](1 + HMD_i \times \theta_7)(1 + IND1 \times \theta_8) \\
&\quad (1 + IND2 \times \theta_9)(1 + BPD \times \theta_{10}) \\
\lambda_{i,j} &= \frac{PRE \times K_{\mathrm{in}}}{(K_{\mathrm{in}} - K_{\mathrm{out},i})}\left(e^{-K_{\mathrm{out},i} \times PNA_{i,j} \times 7} - e^{-K_{\mathrm{in}} \times PNA_{i,j} \times 7}\right)e^{\eta_i} \\
Y_{i,j} &= \frac{\lambda_{i,j}^{y_{i,j}}\,e^{-\lambda_{i,j}}}{y_{i,j}!}
\end{aligned}
\tag{27.17}
$$

where $GA - 24$ is gestational age centered on 24 weeks, $IND1$ assumes the value 1 when RACE is African-American and 0 otherwise, and $IND2$ assumes the value 1 when RACE is Hispanic and 0 otherwise. All other variables are as previously defined.

27.4.4   Exposure–Response Model

With completion of the time course and covariate components of the model, focus turned to determining a model to describe the influence of theophylline on apnea frequency. For this analysis the exposure metric was an approximate average steady-state concentration ($C_{\mathrm{avg}}^{\mathrm{ss}}$). The general form of the exposure–response model was

$$
\lambda_{i,j} = BASE_{i,j} - BASE_{i,j} \times MXRD \times INH_{i,j}
\tag{27.18}
$$

where $BASE_{i,j}$ represents what was defined as $\lambda_{i,j}$ in Eq. (27.16), $MXRD$ is the maximum achievable reduction with treatment, and $INH_{i,j}$ is the inhibition function that describes the fraction of the maximum reduction as a function of theophylline

concentration. The combination of *MXRD* and *INH* is a typical E_{max} model. Models were considered that either fixed the estimate of *MXRD* at 1 or estimated the parameter. The inhibition function (Eq. (27.19a)) and a sigmoidal inhibition function (Eq. (27.19b)) were both evaluated. However, convergence of the estimation algorithm could not be achieved when the Hill coefficient (γ) was included.

$$INH_{i,j} = \frac{C_{avg}^{ss}}{C_{avg}^{ss} + IC_{50}} \tag{27.19a}$$

$$INH_{i,j} = \frac{C_{avg}^{ss\ \gamma}}{C_{avg}^{ss\ \gamma} + IC_{50}^{\gamma}} \tag{27.19b}$$

The model with the *MXRD* parameter estimated led to a larger decrease in the objective function value (-77) versus the model with *MXRD* fixed at a value of 1 (-67); therefore, the former model was selected. The model was unable to support inclusion of an interindividual variance term on IC_{50}.

The fully parameterized model elucidated thus far was subject to the backward elimination procedure to achieve parsimony. When a parameter was removed from the model, an objective function value increase of at least 7.88, corresponding to a nominal significance level of 0.005, was required for retention of the covariate relationship quantified by the parameter. Ultimately, the effect of RACE on K_{out} was removed, as was the effect of gestational age on *PRE* and K_{in}. The influence of bronchopulmonary dysplasia was also determined to be insignificant. Removal of these covariates resulted in the final PD model as follows:

$$
\begin{aligned}
PRE &= \theta_1 \\
K_{in} &= \theta_2 \\
K_{out,i} &= [\theta_3 + \theta_4 (GA_i - 24)](1 + HMD_i \times \theta_5) \\
TVBASE_{i,j} &= \frac{PRE \times K_{in}}{(K_{in} - K_{out,i})} \left(e^{-K_{out,i} \times PNA_{i,j} \times 7} - e^{-K_{in} \times PNA_{i,j} \times 7} \right) \\
BASE_{i,j} &= TVBASE_{i,j} \times e^{\eta_1} \\
IC50 &= \theta_6 \\
MXRD &= \theta_7 \\
INH &= \frac{C_{Theo}}{C_{Theo} + IC50} \\
\lambda_{i,j} &= BASE_{i,j} - BASE_{i,j} \times MXRD \times INH \\
Y_{i,j} &= \frac{\lambda_{i,j}^{y_{i,j}} e^{-\lambda_{i,j}}}{y_{i,j}!}
\end{aligned}
\tag{27.20}
$$

where *PRE*, K_{in} and K_{out} parameterize the model for the time course of apneic episode frequency, *GA* is gestational age in weeks, *HMD* assumes the value of 1 for a diagnosis of hyaline membrane disease and 0 otherwise, *TVBASE* is the expected value of baseline spell count on each postnatal day, postnatal age in weeks (*PNA*) is converted to days with multiplication by 7, η_1 is a random

variable parameterizing the unexplained intersubject variability in the baseline count frequency (*BASE*), and IC_{50} is the concentration of theophylline producing 50% of the theoretical maximum reduction in daily spell count (*MXRD*). The function for *Y*, the predicted probability of the observed spell count, is a Poisson distribution function with mean λ.

The final parameter estimates, relative standard error, and significance level based on the likelihood ratio test for the final PD model are contained in Table 27.1. The fixed effects parameters for *PRE* and K_{in} could not be tested without dismantling the time course model. The statistical advantage of this model was definitively established during model elaboration. IC_{50} was not tested for statistical significance, as its removal from the model disassembles the drug effect model. The significance of the drug effect model was assessed through testing *MXRD*. All retained parameters were highly statistically significant based on the likelihood ratio test. The percent relative standard error (%RSE) was calculated by dividing the asymptotic standard errors by the final parameter estimate and converting to a percentage. The %RSE is a measure of the precision with which the parameter is estimated. The final model suggests that the IC_{50} was 4.26 µg/mL; however, the parameter was imprecisely estimated (77% RSE). The maximum achievable reduction was estimated at 58%. The complete NONMEM control file is contained in the appendix (Model 3).

The estimated probability and conditional probability for episode frequency are contained in Figures 27.5 and 27.6. The estimated probability incorporates all time course and covariate data. The conditional probability also includes the subject level realization of η. Each figure also shows the average probability for each spell count. It is apparent by examining the average probability that the final model performs considerably better than the initial Poisson model in capturing the frequency of the observed data. Note also that the model accounts for zero-inflation and overdispersion. Figures 27.7 and 27.8 plot the individualized (or conditional) prediction of daily spell count versus the observed daily spell count for the base and final model, respectively. The improved distribution of points about the line of identity suggests better performance of the final model when compared to the base model. However, significant deficiency of the model in predicting the highest spell count days is evident. Additionally, the individualized predictions are not "tightly"

TABLE 27.1   Final PD Model Parameter Estimates, Standard Errors, and Level of Statistical Significance[a]

Parameter	PD Parameter	Estimate	%RSE	Statistical Significance
θ_1	*PRE*	5.32	18.1	NT
θ_2	K_{in}	0.28	18.6	NT
θ_3	K_{out}~intercept	0.0333	17.1	$P \ll 0.001$
θ_4	K_{out}~(GA − 24)	0.00402	34.8	$P \ll 0.001$
θ_5	K_{out}~HMD	−0.236	38.2	$P \ll 0.001$
θ_6	IC_{50}	4.26	76.8	NT
θ_7	*MXRD*	0.582	26.5	$P \ll 0.001$
ω_1^2	η_1	0.723	13.6	NT

[a] The estimates are the maximum likelihood estimates determined by NONMEM. %RSE is the percent relative error calculated by dividing the asymptotic standard error by the parameter estimate. Statistical significance is the significance level as determined by the log likelihood difference. NT = not tested.

FIGURE 27.5   Final PD model predicted probability of daily episode counts. The individual points represent the final PD model predicted probability of observing the respective spell counts for every patient day ($n = 4446$). The probability is based on the population estimate of the Poisson mean λ, given subject level covariates, postnatal age, and theophylline concentration. The line is an average probability across all subjects by postnatal day.

FIGURE 27.6   Final PD model predicted conditional probability of daily episode counts. The individual points represent the final PD model predicted conditional probability of observing the respective spell counts for every patient day ($n = 4446$). The conditional probability is based on the individualized estimate of the Poisson mean λ, given subject level covariates, postnatal age, and theophylline concentration, and the subject level realization of η, the interindividual random effect. The line is an average probability across all subjects and postnatal days.

FIGURE 27.7 Subject level predicted versus observed daily apneic episode counts, base model. Predicted number of apneic episodes per day versus observed apneic episodes per day. The line is the line of identity. The predictions are based on individual estimates of λ, the Poisson mean. The predictions are individualized as they arise from the subject level realization of η, the interindividual random effect.

FIGURE 27.8 Subject level predicted versus observed daily apneic episode counts, final model. Predicted number of apneic episodes per day versus observed apneic episodes per day. The line is the line of identity. The predictions are based on individual estimates of the Poisson mean λ, given subject level covariates, postnatal age, theophylline concentration, and the subject level realization of η, the interindividual random effect.

scattered about the line of identity. This was attributed to the substantial within-subject variability. A direct comparison of the performance of the base and final models relative to observed data is presented in Figure 27.9. The average number of daily episodes observed, and conditionally predicted by the base and final models, was calculated across neonates on each postnatal day. The superior performance of the final model in describing the time course of apneic episode frequency is readily apparent. The initial rise in frequency and peak at 1–1.5 postnatal weeks is well characterized. The subsequent resolution is also effectively captured. It should be noted that the points after postnatal week 16 arise from one individual.

The use of a biexponential equation with postnatal age as the time scale permits some practical interpretation of the time course component of the final PD model. Table 27.2 presents the peak spell frequency, the time to achieve peak frequency, and the model predicted resolution half-time of apnea in absence of therapy. The resolution half-time defines the number of days of postnatal maturation that transpire before the daily spell frequency is reduced by one-half. The influence of hyaline membrane disease on resolution half-time is readily apparent. The most premature neonates with HMD have the slowest time to maximum episode counts and have the highest frequency of apnea. A 24 week gestational age infant with HMD requires an additional 7 days for a maturational reduction in spell count of one-half. The half-time of apnea onset is approximately 2.5 days. On average, the greatest severity of apnea would occur at approximately 1 postnatal week. Figure 27.10 depicts the baseline apneic episode frequency versus postnatal age for each gestational age in the present study. The predictions of daily spell count are population predictions, calculated using the final parameter estimates for PRE, K_{in}, and K_{out}.

FIGURE 27.9 Predicted and observed average daily apneic episode count versus postnatal age. The average daily apneic episode count was calculated as the mean across all neonates for each postnatal day. The observed data are represented by the solid circles, the plus denotes average conditional predictions from the base PD model, and the open circles represent the average conditional predictions from the final PD model.

TABLE 27.2 Maximum Apnea Severity, Time of Maximum Severity, and Resolution Half-time Predicted by Final Model[a]

Gestational Age (weeks)	No HMD			HMD		
	t_{max} (days)	Maximum Spells per Day (#)	Resolution Half-time (days)	t_{max} (days)	Maximum Spells per Day (#)	Resolution Half-time (days)
24	8.6	4.0	20.8	9.4	4.2	27.2
25	8.3	3.9	18.6	9.1	4.1	24.3
26	8.0	3.8	16.8	8.8	4.0	21.9
27	7.8	3.7	15.3	8.5	4.0	20.0
28	7.5	3.7	14.0	8.3	3.9	18.4
29	7.3	3.6	13.0	8.1	3.8	17.0
30	7.1	3.5	12.1	7.9	3.8	15.8
31	6.9	3.5	11.3	7.7	3.7	14.8
32	6.8	3.4	10.6	7.5	3.7	13.9
33	6.6	3.4	10.0	7.3	3.6	13.1
34	6.5	3.3	9.4	7.2	3.6	12.3
35	6.3	3.3	8.9	7.0	3.5	11.7

[a] The predicted time to maximum severity (t_{max}), the expected maximum number of spells per day, and the resolution half-time were calculated from the final PD model parameter estimates. These figures represent the anticipated disease severity and time course in the absence of theophylline therapy. HMD = hyaline membrane disease.

FIGURE 27.10 Model predicted baseline apneic episode frequency. The population prediction of the baseline daily apneic episode counts in the absence of drug effect. The population estimates are calculated using the final PD model parameter estimates in the biexponential function describing the time course of neonatal apnea. Each line represents a gestational age: the top line represents a 24 week gestation infant, the bottom line a 35 week gestation infant.

This example provides one approach to the population PK/PD analysis of count data. The reader is directed to the research of Gupta et al. (6) regarding the application of these methods to the analysis of the effect of oxybutinin on urge urinary incontinence and the work of Miller and colleagues (1, 14) on the PK/PD of pregabalin in refractory partial seizures. Cox and colleagues (4) take an alternative, but highly informative, approach by viewing count data as repeated measures time-to-event data in the evaluation of the antiemetic effect of ondansetron.

27.5 SUMMARY

This chapter endeavors to show that a population PK/PD approach to the analysis of count data can be a valuable addition to the pharmacometrician's toolkit. Nonlinear mixed effects modeling does not need to be relegated to the analysis of continuously valued variables only. The opportunity to integrate disease progression, subject level covariates, and exposure–response models in the analysis of count data provides an important foundation for understanding and quantifying drug effect. Such parametric models are invaluable as input into clinical trial and development path simulation projects.

REFERENCES

1. R. Miller, B. Frame, B. Corrigan, P. Burger, H. Bockbrader, E. Garofalo, and R. Lalonde, Exposure–response analysis of pregabalin add-on treatment of patients with refractory partial seizures. *Clin Pharmacol Ther* **73**:491–505 (2003).

2. P. Thall and S. Vail, Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**:657–671 (1990).

3. C. Godfrey, *The Population Pharmacokinetics and Pharmacodynamics of Theophylline in Neonates with Apnea of Prematurity*. PhD Dissertation, University of Connecticut, 2001.

4. E. Cox, C. Veyrat-Follet, S. Beal, E. Fuseau, S. Kenkare, and L. Sheiner, A population pharmacokinetic–pharmacodynamic analysis of repeated measures time-to-event pharmacodynamic responses: the antiemetic effect of ondansetron. *J Pharmacokinet Pharmacodyn* **27**:625–644 (1999).

5. S. Paul, X. Jiang, S. Rai, and U. Balasooriya, Test of treatment effect in pre-drug and post-drug count data with zero-inflation. *Statist Med* **23**:1541–1554 (2004).

6. S. Gupta, G. Sathyan, E. Lindemulder, P. Ho, L. Sheiner, and L. Aarons, Quantitative characterization of therapeutic index: application of mixed-effects modeling to evaluate oxybutynin dose–efficacy and dose–side effect relationships. *Clin Pharmacol Ther* **65**:672–684 (1999).

7. J. DeCosse, H. Miller, and M. Lesser, Effect of wheat fiber and vitamins C and E on rectal polyps in patients with familial adenomatous polyposis. *J Natl Cancer Inst* **81**:1290–1297 (1989).

8. T. Stukel, Comparison of methods for the analysis of longitudinal interval count data. *Statist Med* **12**:1339–1351 (1993).

9. M. Karlsson, A. Thomson, E. McGovern, P. Chow, T. Evans, and A. Kelman, Population pharmacokinetics of rectal theophylline in neonates. *Ther Drug Monit* **13**:195–200 (1991).

10. T. Lee, B. Charles, P. Steer, V. Flenady, and T. Grant, Theophylline population phar-macokinetics from routine monitoring data in very premature infants with apnoea. *Br J Clin Pharmacol* **41**:191–200 (1996).

11. E. Moore, R. Gaix, R. Banagale, and T. Grasela, The population pharmacokinetics of theophylline in neonates and young infants. *J Pharmacokinet Pharmacodyn* **17**:47–66 (1989).

12. G. Casela and R. Berger, *Statistical Inference*. Duxbury Press, Belmont, CA, 1990.

13. P. Thall, Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**:197–209 (1988).

14. B. Frame, R. Miller, and R. Lalonde, Evaluation of mixture modeling with count data using NONMEM. *J Pharmacokinet Pharmacodyn* **30**:167–183 (2003).

15. D. Lambert, Zero-inflated Poisson regression, with an application to defects in manufac-turing. *Technometrics* **34**:1–14 (1992).

16. M. Dalrymple, I. Hudson, and R. Ford, Finite mixture, zero-inflated Poisson, and Hurdle models with application to SIDS. *Comput Statist Data Anal* **41**:491–504 (2003).

17. K. Yau and A. Lee, Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention program. *Stat Med* **20**:2907–2920 (2001).

18. K. Wang, K. Yau, and A. Lee, A zero-inflated Poisson mixed model to analyze diagno-sis related groups with majority of same-day hospital stays. *Comput Methods Programs Biomed* **68**:195–203 (2002).

APPENDIX 27.1

Example of Apnea Data Set

ID	EVID	CURRWT	PNA	PCA	GA	HMD	THEO	DV	FACT
1	0	1.238	0.143	30.143	30	0	5.0099	3	6
1	0	1.181	0.286	30.286	30	0	6.2369	0	1
1	0	1.16	0.429	30.429	30	0	7.3382	0	1
1	0	1.168	0.571	30.571	30	0	7.63	0	1
1	0	1.209	0.714	30.714	30	0	7.4472	0	1
1	0	1.182	0.857	30.857	30	0	7.6099	5	120
1	0	1.205	1	31	30	0	6.5947	3	6
1	0	1.256	1.143	31.143	30	0	8.3511	3	6
1	0	1.319	1.286	31.286	30	0	7.6242	4	24
1	0	1.364	1.429	31.429	30	0	7.1237	4	24
1	0	1.407	1.571	31.571	30	0	8.0082	5	120
1	0	1.465	1.714	31.714	30	0	7.5126	2	2
1	0	1.493	1.857	31.857	30	0	7.2322	5	120
1	0	1.519	2	32	30	0	6.969	2	2
1	0	1.564	2.143	32.143	30	0	6.9914	1	1
1	0	1.594	2.286	32.286	30	0	7.0676	1	1
1	0	1.673	2.429	32.429	30	0	6.8272	2	2

NONMEM Models

Model 1

```
$PROB RUN# 001 Poisson Distribution
$INPUT C,ID,DV,FACT,. . .  ;other variables as needed
$DATA data.csv
$PRED
;Log canonical link used for example
TLLM=THETA(1)   ;Typical value of log(lambda)
LLM=TLLM+ETA(1) ;Log(lambda) with interindividual variability
LM=DEXP(LLM)     ;Lambda with IIV
TLM=DEXP(TLLM)  ;Population Lambda (For table)

;Poisson Distribution
;The  DV  factorial  is  supplied  via  the  dataset  as  FACT
POIS=(LM**DV)*DEXP(-LM)/FACT
;Typical value of Poisson
;(coded for inclusion in table)
;(gives probability based on population mean)
TPOIS=(TLM**DV)*DEXP(-TLM)/FACT
Y=POIS
$THETA (0,1,5)
$OMEGA 0.01
$ESTM NOABORT PRINT=5 MAXEVAL=9999 METHOD=1 LAPLACE LIKELIHOOD
$TABLE ID TLLM LLM LM TLM POIS TPOIS DV Y ONEHEADER NOPRINT
FILE=001.TAB
```

Model 2

```
$PROB RUN# 002 Zero-Inflated Poisson (ZIP)

$INPUT C,ID,DV,FACT,. . .  ;other variables as needed
$DATA data.csv
$PRED
;Model for the state: susceptible (1) vs non-susceptible (0)
;Logit canonical link
TLOGIT=THETA(1) ;typical value of logit
TPHI=DEXP(TLOGIT)/(1+DEXP(TLOGIT)) ;typical value of probability
LOGIT=TLOGIT + ETA(1) ;Logit with IIV
PHI=DEXP(LOGIT)/(1+DEXP(LOGIT)) ;Individual Probability

;Given susceptible state, probability of count
; Log canonical link used for example
TLLM=THETA(1)   ;Typical value of log(lambda)
LLM=TLLM+ETA(1) ;Log(lambda)
LM=DEXP(LLM)     ;Lambda with IIV
TLM=DEXP(TLLM)  ;Population Lambda (For table)
;Poisson Distribution
```

```
;The DV factorial is supplied via the dataset as FACT
POIS=(LM**DV)*DEXP(-LM)/FACT
TPOIS=(TLM**DV)*DEXP(-TLM)/FACT


STATE=0
IF (DV.EQ.0) STATE=1


P0=PHI + (1-PHI)*DEXP(-LM) ;Probability of zero count
PN=(1-PHI)*POIS ;Probability of count 1,2,3,...
ZIP=P0**STATE * PN**(1-STATE) ;
Y=ZIP
;for table
EYI=(1-PHI)*LM          ;Subject-level expectation
VYI=EYI+EYI*(LM-EYI)  ;Subject-level variance of count
EYP=(1-PPHI)*TLM        ;Population expectation
VYP=EYP+EYP*(TLM-EYP) ;Population variance of count


$THETA (-8,-0.6,8) (0,2.2)
$OMEGA 0.3 0.4
$ESTM NOABORT PRINT=5 MAXEVAL=9999 METHOD=1 LAPLACE LIKELIHOOD
$TABLE ID DV TLLM LLM TLM LM TPOIS POIS LOGIT PPHI PHI
P0 PN DV EYI VYI EYP VYP DVY ETA1 ETA2
ONEHEADER NOPRINT FILE=002.TAB
```

Model 3

```
$PROB RUN# Final PD model
$INPUT C,ID,EVID,WT,PNA,PCA,GA,HMD,THEO,DV,FACT
$DATA data.csv
$PRED
;Time course and covariate model
TPRE=THETA(1) ;
TIN=THETA(2)
OUT=THETA(3)+THETA(4)*(GA-23.9)
TOUT=OUT*(1+THETA(5)*HMD)
TEX1=DEXP(-TIN*(PNA*7))
TEX2=DEXP(-TOUT*(PNA*7))
TBAS=(TPRE*TIN)/(TIN-TOUT)*(TEX2-TEX1)
;-------------
;Interindividual variability model
BASE=TBAS*DEXP(ETA(1))
;-------------
;Exposure Response Model (ERM)
IC50=THETA(6) ;IC50 for theophylline conc
MXRD=THETA(7) ;Maximal Reduction possible
INH=THEO/(THEO+IC50)
ERM=BASE*MXRD*INH
LM=BASE-ERM
```

```
;models the likelihood
;f(y)=Poisson=(lambda**y)*exp(-lambda)/y!
T1=LM**DV
T2=DEXP(-LM)
POIS=T1*T2/FACT
Y=POIS
$THETA (0,5.3)(0,0.28)(0,0.033)(-2,0.004,5)(-2,-0.239,2)
(0,4.28)(0,.584,1)
$OMEGA 0.725
$ESTM NOABORT PRINT=5 MAXEVAL=9999 METHOD=1 LAPLACE LIKELIHOOD
$COVARIANCE
$TABLE ID THEO PNA GA HMD TPRE TIN TOUT DV TBAS
BASE IC50 MXRD INH ERM ETA1 LM Y ONEHEADER NOPRINT FILE=final.TAB
```

Mixture Modeling with NONMEM V

BILL FRAME

28.1 INTRODUCTION/MOTIVATING EXAMPLES

Examples exist in pharmacology where either the pharmacokinetics or the pharmacodynamics of a drug is polymorphic within a population. Several examples exist involving acetylation polymorphism. Slow acetylators of isoniazid have a higher incidence of peripheral neuropathy secondary to elevated concentrations, as well as an increased likelihood of hepatotoxicity (1). In the isoniazid example, a pharmacokinetic (PK) analysis of a data set without acetylator status as a covariate would likely demonstrate drug clearance to be distributed bimodally. An additional dimension of complexity for the isoniazid acetylation example arises as follows. The acetylation is under monogenic control, whereby slow acetylators are homozygous for a recessive allele pair. The partition between slow and fast acetylators is roughly 50%/50% in Caucasians and blacks, yet in Orientals and Eskimos the partition is closer to 15% with slow, and 85% with fast acetylation (1). In this example, if one had the covariate associated with the population polymorphism in the data set, then the observations would be more easily explained by including the covariate in the model (i.e., include acetylator genotype as a covariate on isoniazid clearance).

Polymorphic expression of alleles, enzymes, or drug transporters has been associated with variations in response to several classes of drugs (1–15). Unfortunately, situations arise where there is some type of polymorphism in the response being modeled, and even though there may be a covariate that explains the polymorphism, the covariate may not be available to the modeler. This leads to the utility of a mixture model. Mixture modeling involves modeling a probability distribution with a mixture of probability distributions. Other terms loosely synonymous with mixture modeling include unsupervised learning, latent variable analysis, and clustering. A mixture model allows each individual's data to be described by two or more different models—hence the designation "mixture." With a mixture model it is assumed that the population is partitioned into two or more subpopulations according to some probability model, and that each subpopulation has its own submodel, which differs from the other submodels with respect to fixed or random

effect parameters. The probability model may be relatively simple (i.e., there is one fixed probability partition for the population) or it may depend on several covariates (i.e., the partition may shift dynamically as covariate values change between subjects) and is referred to as a dynamic mixture. One might also assume that one probability partition describes a polymorphism across several model parameters (i.e., the same probability partition holds for volume of distribution and for the clearance of a drug), or that each model parameter that is believed to be polymorphically distributed has its own probability partition (i.e., there could be a 20–80% partition on clearance but a 50–50% partition on volume of distribution). This last situation is called a multiple mixture.

The previous isoniazid example suggested using a mixture model in a situation where there was a biological explanation for the mixture but where an explanatory covariate was missing. These are called biological mixtures. Another possible use for mixture models is the situation where a random effect has a skewed distribution, which cannot be made symmetric through modification of the structural model or transformation of the random effect. Here, a mixture of symmetric distributions for the random effect may better approximate the skewed distribution than a single distribution. Mixtures used in this fashion are called statistical mixtures. In the remaining sections of this chapter these concepts and others as they relate to mixture modeling using NONMEM V are explored. A passing familiarity with NONMEM parlance is helpful but not necessary to digest this chapter. Those without previous NONMEM experience are referred to *NONMEM Users Guides* five, seven, and eight, which serve as a nice introduction to the use of NONMEM (16). Numerous NONMEM control stream examples illustrating the concepts presented in this chapter are provided, so the reader can learn by running and modifying the NONMEM code, which can be found on the book's ftp site. While the examples all involve simulated data, the principles demonstrated can be applied to drug development or therapeutic drug monitoring. The author freely acknowledges that almost all the concepts presented herein have been previously described in the NONMEM documentation.

28.2   HISTORY

In 1894 Karl Pearson (17) proposed a method of decomposing a mixture of two univariate distributions with unequal variances. Pearson's method required solving a ninth degree polynomial. During the 1930s and 1940s a graphical approach to mixture decomposition was popular (18). Likelihood estimation of mixture parameters, which required computing machinery, was first suggested by Rao (19). While several textbooks have been written about mixture modeling from a statistical perspective, the pharmacostatistical literature is sparse in terms of examples of mixture modeling, with most of the work appearing in statistics, genetics, agriculture, or psychology journals (20–29).

The published work is divided between papers addressing applications, parameter estimation methods, and tests of hypotheses regarding mixture existence. Two NONMEM related papers described two subpopulations with respect to the clearance of intravenously administered cephalosporins (20, 21). The antianginal agent perhexilene, which is known to have polymorphic cytochrome P450-2D6

metabolism, was estimated to have two subpopulations with respect to clearance and volume of distribution (23). Simulations were used to compare parameter estimation from parallel, crossover, and dose escalation designs in the presence of a subpopulation of nonresponders (24). Other simulation exercises have explored the distribution of the differences in minimum objective function values between mixture and nonmixture analysis of both mixed and nonmixed subpopulations to assess the probability of finding a mixture when one is not present, and the probability of finding a mixture and classifying its subjects correctly when it truly does exist (29). In another study, subpopulations were identified with respect to sensitivity to stroke induced sedation while exploring the use of chlormethiazole as a neuroprotective agent (25).

Note All •TXT files referred to in the subsequent sections can be found in the book's ftp site.

28.3 SUBMODEL PARAMETERIZATIONS

As stated in the introduction, the submodels may differ in fixed or random effects. The task is to learn how to communicate our ideas about the submodels to NONMEM. Suppose that during a population pharmacokinetic (PK) analysis of an orally administered drug, a model is used where the absorption and elimination rates are first order. The model is parameterized in terms of elimination rate (K), apparent volume of distribution (Vd) and absorption rate (KA), such that both K and KA are allowed to vary between subjects. Specifically, the values of K and KA for the jth subject (K_j and KA_j) are specified as follows:

$$K_j = THETA(1) \cdot e^{\eta_{1j}} \tag{28.1}$$

$$KA_j = THETA(2) \cdot e^{\eta_{2j}} \tag{28.2}$$

Here, the random effects η_{ij}, which allow for between-subject variability in the PK parameters, are assumed to be normally distributed about zero, with variance Ω (i.e., $\eta_{ij} \sim N(0, \Omega)$, and Ω is called a random effects parameter). The apparent volume of distribution for the jth subject is modeled as a linear function of weight (WT_j) without any random effect.

$$Vd_j = THETA(3) \cdot WT_j \tag{28.3}$$

The residual, or intrasubject, error is being modeled with a proportional error model as follows:

$$C_{ij} = C_{mij} \cdot (1 + \varepsilon_{ij}) \tag{28.4}$$

where C_{ij} (C_{mij}) is the ith observed (predicted) concentration for the jth individual, and ε_{ij} is a random effect assumed to be normally distributed with zero mean and variance Σ. Finally, suppose there are 100 subjects, each with 24 observed concentrations, and the method of estimation being used is the first-order conditional method

with ε–η interaction. The file naming system for this chapter uses CZ.TXT, RZ.TXT, TZ.TXT, and TZA.TXT, for the control stream, report output file, NONMEM output table, and FIRSTONLY NONMEM table, respectively (where Z indexes the model sequence). A plot of the density of the modal estimates of the random effect on the elimination rate constant reveals positive skewness (Figure 28.1).

Ideally, the skewness coefficient is zero. The presence of skewness in this random effect plot is undesirable and data simulated with the model might not statistically reflect the observed data. The absence of bimodality in the plot does not imply the absence of a mixture. Indeed, when a mixture is present, and the means of two subpopulations are close together, or their variances are large, or their relative proportions are unbalanced (i.e., a 90%/10% partition versus a 50%/50% partition), one may not see bimodality in these types of plots. Two possible approaches to this skewness problem are mixture modeling and random effect transformation.

First, consider a two subpopulation mixture model. The $PK block of the control stream needs to be modified to communicate to NONMEM that there may be two subpopulations with respect to the elimination rate constant, K. A variable called MIXNUM indexes the subpopulation, and hence the submodel, for which variables are computed. MIXNUM may be used as a right-hand quantity in the abbreviated code, as long as the control stream contains a special block of abbreviated code called $MIX (see Section 28.4). For the example, the code for K is modified in the $PK block as follows:

```
$PK
IF (MIXNUM.EQ.1) THEN
  K=THETA(2)*EXP(ETA(2)) ; K for first subpopulation
ELSE
  K=THETA(4)*EXP(ETA(3)) ; K for second subpopulation
ENDIF
```


FIGURE 28.1   A density plot superimposed on a histogram of modal values of random effect on elimination rate for one subpopulation: conditional estimation with ε–η interaction.

(Note that the rest of the code is still needed to run NONMEM successfully, but the parts of the code relevant to the mixture model are presented in this chapter.)

This code instructs NONMEM to allow K to have the value of THETA(2) for subpopulation one (MIXNUM=1) and THETA(4) for subpopulation two (MIXNUM=2). It also appears that the two subpopulations might differ in their random effects as observed from the fact that different random effects were introduced for each subpopulation (ETA(2) and ETA(3)). For this first example, it is assumed that ETA(2) and ETA(3) have the same variance (Ω), so that there is really only one random effects parameter for *K*, and it is shared between the two subpopulations. To accomplish this the $OMEGA block is modified:

```
$OMEGA .25            ; random effects parameter associated with KA
$OMEGA BLOCK(1) .15   ; random effects parameter associated with K
                        for subpopulation one
$OMEGA BLOCK(1) SAME  ; random effects parameter associated with K
                        for subpopulation two
```

Before moving onto describing the probability model (the $MIX block), a couple of comments are in order. The first involves what might be done differently with the coding of the random effects part of the model. Had the following $PK and $OMEGA code been used,

```
$PK
IF (MIXNUM.EQ.1) THEN
  K=THETA(2)*EXP(ETA(2))
ELSE
  K=THETA(4)*EXP(ETA(2))
ENDIF
$OMEGA .25            ; random effects parameter for KA
$OMEGA BLOCK(1) .15  ; random effects parameter for K for subpopulation
                        one
```

the same results as that produced by the first code would have been obtained. It might seem odd that the same random effect is used for an individual, regardless of which subpopulation that individual is in, especially with conditional estimation, where modal estimates of these random effects are computed as part of the minimization procedure. When this type of parameterization is encountered by NONMEM, the second subpopulation is given its own separate random effect (like the ETA(3) in the first parameterization). The marginal likelihood for the *i*th subjects data, L_i, is given by

$$L_i = \sum_{k=1}^{r} p_{ik}(\theta) \cdot \int l_{ik}(\eta, \theta) \cdot h(\eta, \Omega) d\eta \tag{28.5}$$

Equation (28.5) gives the average of the marginal likelihoods over the *r* subpopulations. Thus, it can be seen how the mixture probability ($p_{ik}(\theta)$) can vary between subpopulations (as *k* varies) and between individuals (as *i* varies) due, for example,

to covariate variation between subjects. This equation also suggests that for the same subject, the variable of integration, η, can assume different modal values for the different subpopulations. The term $h(\eta, \Omega)$ is the density function for η and the term $l_{ik}(\eta, \theta)$ is the conditional likelihood. For each subpopulation, the empirical posterior probability that a subject's data is described by the kth submodel is given by the following expression:

$$p_{ik}(\theta) \cdot \int l_{ik}(\eta, \theta) \cdot h(\eta, \Omega) d\eta / L_i \qquad (28.6)$$

For each subject the value of k corresponding to the largest of these r values has the name MIXEST within NONMEM abbreviated code.

Second, it is not necessary that ETA(2) and ETA(3) should have the same variances. It could be assumed that these variances were different with the following $OMEGA block.

```
$OMEGA .25            ; random effects parameter associated with KA
$OMEGA BLOCK(1) .15 ; random effects parameter associated with K
                       for subpopulation one
$OMEGA BLOCK(1) .25 ; random effects parameter associated with K
                       for subpopulation two
```

Whether one is able to fit mixture models with distinct random effects parameters for each subpopulation is dependent on the nature of the underlying mixture. Are the subpopulations close together in mean, how much data is available (per subject and total), and which type of estimation is being used (first order, hybrid, Laplacian)? Now to complete the attempt at applying a two subpopulation mixture model to this data, the probability model and number of subpopulations must be communicated to NONMEM via the $MIX block. Within the $MIX abbreviated code the number of subpopulations are communicated with the variable NSPOP and the probabilities associated with the subpopulations with the variable P(i) (or its alias MIXP(i)), where i indexes the subpopulation. Thus, the code would be

```
$MIX
NSPOP=2        ; there are two subpopulations
P(1)=THETA(5) ; the probability of being in the first subpopulation
                 is THETA(5)
P(2)=1-P(1)   ; the probability of being in the second subpopula-
                 tion is 1-THETA(5)
```

Since THETA(5) is a probability it must be constrained to the interval [0, 1] in the $THETA block. Also note that changing from the nonmixture model to the mixture model required the addition of two new THETA parameters. One was used to control the probability partition, and the other to specify how the two subpopulations differed. For now, note that neither can be entered into the model uniquely. They must both go into the model together, or be removed from the model together (the designated driver system), and this leads to issues regarding the hypothesis testing for the presence of a mixture (see Section 28.5). Two control stream/report/output table pairs (C2.TXT/R2.TXT/T2A.TXT and C3.TXT/R3.TXT/T3A.TXT) can be

found in the book's ftp site. They differ only in how the ETA random effects are coded as discussed previously. Neither one concluded with a successful $COVARIANCE step and they both have similar minimum objective function (MOF) and parameter estimate values. It is noteworthy that the final MOF for these runs (−8232.5) is lower than that obtained with the nonmixture model (−8208.5) and that it is estimated that the population is partitioned with 94% of the patients having a *K* of 0.11/h and 6% having a *K* of 0.26/h. This distribution of the *K* values could help explain the skewed distribution seen with the nonmixture model.

The example above used two ways of specifying the random effects on *K*, paving the way to a discussion about how NONMEM calculates the eta-bar statistic when conditional estimation is used. Notice that both control streams (C2.TXT and C3.TXT) contain the statement EST=MIXEST in their $PK blocks. At finalization of the NONMEM run, EST will contain the number (1 or 2) corresponding to each subject's most likely subpopulation. This allows the tabulation of EST and the modal ETA values for each patient using the $TABLE record with the FIRSTONLY option. Each patient will have an ETA estimate, only for his/her most likely subpopulation. His/her ETA estimate for the other subpopulation will be zero. For the first mixture model report (R2.TXT), the output for the eta-bar section is as follows:

```
ETABAR: -0.36E-01 0.49E-03

P VAL.: 0.38E+00 0.98E+00
```

and for the second mixture model attempt (R3.TXT),

```
ETABAR: -0.28E-01 -0.39E-02 -0.50E-02

P VAL.: 0.49E+00 0.87E+00 0.95E+00
```

Disregarding the discrepancy for the *KA* eta-bar values, it can be seen that one value (0.00049) results for our first attempt, where the same ETA was coded for each of the two *K*s. This is the average of the ETAs across all subjects, regardless of subpopulation assignment. For the alternate coding of the ETAs (i.e., 2 ETAs for the subpopulations) two values are returned for ETABAR, as expected. The first is the average of those patients estimated to be in the first subpopulation, and the second is the average for the remaining patients. Ideally, the values obtained for ETABAR for the second alternative can be obtained from the output from the first alternative coding of the ETAs using EST, and then calculating the subpopulation averages. This is not the case here as the ETABAR for R2 is positive and both ETABARs for R3 are negative. It is left as an exercise for the reader to examine the modal ETA estimates and rationalize this apparent contradiction. Neither way is really wrong, as all the ETAs by definition come from the same distribution. It will be seen later that things may not always be this clear-cut.

Now for a final swing at this data with a mixture model, but this time using the first-order method (the default or METH=0 on the $ESTIMATION record). Recalling that skewness in a modal K ETA plot from the original nonmixture model prompted the consideration of a mixture on K, the $ESTIMATION records on both C1.TXT (becomes C5.TXT) and C3.TXT (becomes C4.TXT) are modified as follows:

```
$ESTIMATION MAXEVALS=9000 PRINT=1 POSTHOC;
```

Inclusion of the POSTHOC option instructs NONMEM to obtain the Bayesian post hoc ETA estimates when the first-order method is used. These effects and other relevant parameters can be output into a table using the $TABLE record. Thereafter, the distribution of the effects can be characterized, including skewness if present. Both the mixture model and the nonmixture models need to be reestimated with the first-order method, as one cannot compare the MOFs in a meaningful way between models differing only in estimation method. The MOF has dropped 676 points between the nonmixture model (see R5.TXT) and the mixture model (R4.TXT). Furthermore, the mixture model run has now concluded with a successful $COVARIANCE step. A choice has to made whether to make two plots (one for each subpopulation) or one (after all, the ETAs all share the same distribution). The latter approach is shown in Figure 28.2. Similar plots can be generated for each subpopulation.

One final comment is in order regarding the above parameterizations. This code does not prevent THETA(2) and THETA(4) from reversing roles in terms of which is associated with the larger of the two Ks. Of course, when this happens THETA(5), which estimates the proportion of the population with THETA(2)=K, would become 1-THETA(5). Specifying one K as an appropriate fraction of the other will constrain it to always be associated with the desired subpopulation.

The ideas expressed above can be extended up to four subpopulations with $MIX, and it may be tempting to extend the IF − THEN − ELSE structure to a IF − THEN − ELSEIF − ELSE − ENDIF structure as follows:

FIGURE 28.2   Density superimposed on a histogram of post hoc values for random effect (ETA2) on elimination rate for two subpopulations. The first-order method was used for estimation.

```
IF (MIXNUM.EQ.1) THEN
  ;model for first subpopulation
ELSEIF (MIXNUM.EQ.2) THEN
  ;model for second subpopulation
ELSE
  ;model for third subpopulation
ENDIF
```

Although this is acceptable Fortran coding, it is not acceptable abbreviated code with NONMEM V. Instead, a separate IF − THEN − ELSE − ENDIF block needs to be used for each subpopulation, to avoid error messages about nested random effects.

An alternative approach to deal with the skewness seen in the density for the random effect on K is to transform this random effect (η) to a new random effect, ζ, as follows:

$$\zeta = \frac{e^{\eta\lambda} - 1}{\lambda} \tag{28.7}$$

Here, λ is a real transformation parameter to be estimated by NONMEM. This transformation has the following useful properties.

1. It is invertible.
2. It maps real values of η to real values of ζ.
3. It is everywhere differentiable with respect to η.
4. As λ approaches zero, it becomes the identity transformation.
5. It is sometimes able to correct positive or negative skewness.

It is up to the reader to verify properties 1–4, but a comment is needed regarding property 5. The transformation is most effective when there is much data per subject and conditional estimation is being employed. For the example, the implementation is as follows:

```
ET2=(EXP(ETA(2)*THETA(4))-1)/THETA(4)
K=THETA(2)*EXP(ET2)
```

Recalling that one cannot use zero as a starting estimate for the transformation parameter (λ=THETA(4)), the new model (see C6.TXT R6.TXT) has an objective function value that is 16 points lower than the original model that we started with (C1.TXT R1.TXT), similar fixed effects parameter estimates, but substantially reduced skewness in the density of the random effect on K (see Figure 28.3).

One must be careful not to become a lazy modeler and rely on mixtures when covariates, structural model modifications, or changes in random effects structure can be used in lieu of a mixture. Before one accepts a mixture model, the covariates should be examined closely, across the subpopulations, with the idea of seeing a way to include them in a nonmixture model.

FIGURE 28.3   A density plot of modal values of random effect on elimination rate for one subpopulation. The first-order conditional estimation with ε–η interaction (transformed random effect) was used.

28.4   PROBABILITY PARAMETERIZATIONS

The goal of this section is to introduce different ways to communicate probability models to NONMEM. These concepts will be needed in the next two sections. Recall from the first example that it was finally possible to fit a two-component mixture to the data by changing the estimation method from FOCE + INTERAC-TION to the first-order method (see C4.TXT – R4.TXT). The parameterization in $MIX is as follows:

```
$MIX
NSPOP=2 ; there are two subpopulations
P(1)=THETA(5) ; the probability of being in the first is THETA(5)
P(2)=1-P(1) ; the probability of being in the second is 1-THETA(5)
```

The estimate of THETA(5) is 0.92, and its standard error is 0.0271. A 95% Wald based confidence interval can be constructed for the proportion of the population with K=THETA(2) as $0.92 \pm 1.96 \cdot 0.0271$, or [0.87, 0.97]. For this example, the confidence interval contains neither zero nor one, the two boundary points for a probability measure. Had the standard error of THETA(5) been larger, say, 0.1, a confidence interval would include not only one, but an interval above one, which cannot be interpreted as a probability. The ramifications of different probability parameterizations will be explored in Section 28.5. Clearly, a parameterization that constrains all probabilities between zero and one, inclusive, would be ideal. Below is an implementation that almost achieves this, for a three-component mixture.

```
$MIX
  NSPOP=3
  A=EXP(THETA(1))
```

```
B=EXP(THETA(2))
C=1+A+B
P(1)  = A/C
P(2)  = B/C
P(3)  = 1/C
```

The above coding for `P(3)` is preferable to `P(3) = 1 - P(2) - P(3)`, which sometimes generates probabilities that are slightly negative, resulting in NONMEM errors and aborted runs.

28.5 HYPOTHESIS TESTING

This section is intended to introduce some of the issues that arise when one attempts to determine if a mixture truly exists (i.e., there are two or more subpopulations). The discussion is restricted to the case where there are at most two subpopulations. Two main points to consider are: the likelihood of concluding that a mixture of two subpopulations is present when there truly is only one population (false positive significance level); and the ability to identify and quantify a mixture of two subpopulations when it is truly present (power).

The first point is addressed by continuing with the example, letting `P(1)` be the proportion of the population with `K=THETA(2)`, and reparameterizing our mixing probabilities as (see `C7.TXT`, `R7.TXT`)

```
$MIX
  NSPOP=2
  P(1)  = 1/(1+EXP(-THETA(5)))
  P(2)  = 1/(1+EXP(THETA(5)))
```

The null hypothesis H_0, that there is no mixture (i.e., there is only one subpopulation), can be tested against the alternative hypothesis H_a, that there are two unique subpopulations with respect to K. The null hypothesis can be satisfied in either of two ways. If the data is fitted with a two subpopulation mixture model and the two estimates of K are very close together, this indicates that there is no mixture regardless of what `P(1)` is estimated to be. On the other hand, if the possibility exists that `P(1)` is zero or one, then it follows that there is no mixture as this implies that all of the patients are being estimated to be in only one population. From a hypothesis testing perspective, the null hypothesis of no mixture existence (H_0), the union of two sub-null hypotheses ($H_{01} \cup H_{02}$), can be expressed as follows:

```
H₀₁ :P(1)  = 0 or P(1)  = 1  H₀₂  :  |THETA(2)-THETA(4)|< ε
```

for some small $\varepsilon > 0$. To reject H_0 requires rejecting both H_{01} and H_{02}.

The functional form being used for `P(1)` maps the real axis to the open interval $(0, 1)$ and large values for `|THETA(5)|` are associated with `P(1)` values near the boundary points for a probability measure (zero and one). Implicit in this parameterization is that neither `P(1)` nor any points in its confidence interval can be zero or one. This allows subjectivity about how extreme the confidence bounds for `P(1)` need to be to conclude that there is no mixture. Since the boundary points used in H_{01} can never be included in the confidence interval for `P(1)`, we construct a 95%

confidence interval for `THETA(5)` as `THETA(5)±1.96` times its standard error (SE) and arbitrarily consider `P(1)` ≠ 0 and `P(1)` ≠ 1 (i.e., H_{01} is rejected) equivalent to

`(THETA(5)  -1.96·SE(THETA(5)),  THETA(5)  +1.96·SE(THETA(5)))`

being entirely contained in [–3, 3]. This condition is equivalent to [0.048 ≤ 95% confidence bounds for `P(1)` ≤ 0.95]. The constructed confidence interval for `THETA(5)`, [–3.16, –1.72], does not fall within [–3, 3]; therefore H_{01} is not rejected and hence H_0 is not rejected. Had the interval [–4, 4] been arbitrarily selected as the containment interval then H_{01} would have been rejected. Picking a wider inclusion interval would increase both the power and the false positive significance level of the test.

Turning now to H_{02}, are the two values for K indistinguishable? One approach is to construct a 95% confidence interval (CI) for the difference between the two values of K, and if the CI contains zero, do not reject H_{02}. To construct the CI compute `THETA(2) − THETA(4) ± 1.96·SE(THETA(2) - THETA(4))` where `SE(THETA(2) - THETA(4))` is taken to be the positive square root of the quantity variance `(THETA(2)) + variance(THETA(4)) - 2·covariance(THETA(2),THETA(4))`. The CI is [–0.144, –0.108] and does not contain zero, supporting the notion that the two elimination rate constants do differ. An alternative approach to the above would be to replace the Wald based confidence intervals with those produced using the nonparametric bootstrap technique. With this technique the data set is sampled with replacement at the subject level many times, and the model is fit to each of these resampled data sets, generating an empirical distribution for each model parameter. Confidence intervals can then be constructed for the model parameters based on the percentiles of their empirical distributions.

Next, return to the issue of the difference in the `MOF` values between the mixture analysis and nonmixture analysis of a data set. Recall that when a mixture is introduced, two new parameters are required. One might indeed be tempted to compare the difference in `MOF`s between a mixture and nonmixture model to a value from the χ^2 distribution with two degrees of freedom and upper tail area of 0.05 (i.e., 5.99). Unfortunately, there is no theoretical basis for selecting this value, so its choice is arbitrary. Regularity assumptions for hypothesis testing with likelihood ratio tests require that parameter values used in the null hypothesis be in the interior of the parameter space (30). Since both values for `P(1)` in the test H_{01} are on the boundary of the parameter space, the interior assumption is violated.

Considerable work has been focused on determining the asymptotic null distribution of –2 log-likelihood ($-2LL$) when the alternative hypothesis is the presence of two subpopulations. In the case of two univariate densities mixed in an unknown proportion, the distribution of $-2LL$ has been shown to be the same as the distribution of $[\max(0, Y)]^2$, where Y is a standard normal random variable (28). Work with stochastic simulations resulted in the proposal that $-2LL \cdot c$ is distributed χ^2 with d degrees of freedom, where d is equal to two times the difference in the number of parameters between the nonmixture and mixture model (not including parameters used for the probability models) and $c = (n - 1 - p - g/2)/n$ (31). In the expression for c, n is the number of observations, p is the dimensionality of the observation, and g is the number of subpopulations. So for the case of univariate observations ($p = 1$), two subpopulations ($g = 2$), and one parameter distinguishing the mixture submodels (not including the mixing parameter), $-2LL \cdot (n - 3)/n \sim \chi^2$ with two

degrees of freedom, or asymptotically $-2LL \sim \chi^2$ with two degrees of freedom. This result is difficult to interpret in the context of mixed effects modeling, where n might be chosen to be the number of subjects or the number of observations, and p varies from subject to subject (i.e., not all subjects have the same amount of data). Results of stochastic simulation work using NONMEM with seizure count data found that the 95% percentile of the distribution of differences in MOF between mixture and nonmixture analysis of data which was simulated under the null hypothesis of no mixture was 5.43, which is close to 5.99, the 95th percentile for a χ^2 distribution with two degrees of freedom (29). The reader is referred to McLachlan and Basford (27) for a more thorough discussion of the analytic and simulation work that has been done to characterize the distribution of the likelihood ratio test for testing the number of components (subpopulations) in a mixture.

The ability to detect a mixture when it is present can be approached at the subpopulation or subject level. One may focus on how well the submodel or probability model parameters are estimated, or how accurately subjects are classified into subpopulations. For example, mixed populations with submodel parameters that are far apart, partitions that are not highly imbalanced (a 5%/95% partition would be considered imbalanced), or low interindividual variability in the parameter of difference would be easier to characterize than mixed populations not meeting any of these criteria. Accuracy of subject classification can be quantified as the overall proportion of patients assigned to their correct subpopulation, the probability that they are estimated to be in a given subpopulation given that they are truly in that subpopulation (senitivity of estimation), or the probability that they are in a given subpopulation given that they are estimated to be in that subpopulation (predictive value of estimation). These types of power assessments are most easily conducted when the true state of reality is known (29).

28.6   DYNAMIC MIXTURES

Thus far all mixtures considered have been static, in the sense that the probability model did not change as a function of covariate values. Recall that in the introduction, isoniazid acetylator polymorphism was used as an example to introduce the concept of mixture modeling utility. In that example it was stated that race was associated with how patients were partitioned between slow and fast acetylator status. So, given an isoniazid PK data set without acetylator genotype, but with race as a covariate, one might want to introduce race as a covariate in our $MIX block to help model the patients as either fast or slow acetylators.

The example for dynamic mixtures consists of a data set with a total of 900 subjects (DATA2.TXT). The subjects were enrolled in a placebo controlled, parallel group study of an investigational agent for treating compulsive gambling. The subjects were equally distributed between daily dose groups of 0, 1, 2, 3, or 4 mg, without any titration. Collected covariates were limited to age, sex, shoe size, and baseline body weight at screening. The dependent variable for this exercise is body weight in kilograms, which is measured weekly for 12 weeks. Visualization of the data reveals that all 900 subjects completed the study, and in general patients either maintained their baseline body weight, gained weight (monotonically), or lost weight (monotonically).

Based on inspection of the data the following model is implemented using a $PRED block (see C8.TXT R8.TXT):

```
$PRED
W=THETA(1) ; residual error as a standard deviation
AS=THETA(2)+THETA(3)*DOSE+ETA(1) ; individual asymptote as a function
of dose
K=THETA(4) ; rate constant
FA=BSLN*EXP(AS*(1-EXP(-K*TIME))) ; prediction as a function of
baseline, time, and asymptote
Y=FA+W*EPS(1) ; additive residual error model
```

This functional form was chosen as it allows for weight neutrality, gain, or loss. Results from this approach using conditional estimation reveal a poor fit to the data, based on the final gradient (vector of partial derivatives of the objective function with respect to the fixed and random effects parameters) demonstrating several large values. If this fit represents a true global minimum, then it suggests that subjects receiving placebo or 4 mg/day lose about 2% or gain about 8% of their baseline weight, respectively. Inspection of the density of the modal estimates for ETA(1) (Figure 28.4) reveals significant skewness and leptokurtosis. Both of these attributes suggest that a mixture model might be tried, assuming that covariate inclusion attempts had been futile. The spike with support near zero (or atom in statistical parlance) suggests a group with very little deviation from baseline in asymptotic weight. Of course, such a spike could also result from patients with very little data whose ETA(1) estimates might be driven to zero, but there are no such patients.

Next, a three subpopulation mixture model is tried, whereby subjects could remain stable, gain, or lose weight. For this first mixture attempt, drug exposure is not included as a covariate. The three submodels and probability models are as follows (see C9.TXT R9.TXT).

FIGURE 28.4 A density plot of modal values of random effect on weight change asymptote for one subpopulation. The first-order conditional estimation (FOCE) method was used.

```
$PRED
IF (MIXNUM.EQ.1) THEN ; sub model for stable weight
  AS=0 ; typical prediction = baseline
  ET=ETA(1) ; random effect only a place holder, var(ETA(1)==0
  W=THETA(1); residual error as a standard deviation
ENDIF
IF (MIXNUM.EQ.2) THEN ; weight loss
  AS=THETA(6)*EXP(ETA(2)) ; controls asymptotic weight loss
  W=THETA(2) ; residual error as a standard deviation
  K=THETA(4) ; controls rate of weight loss
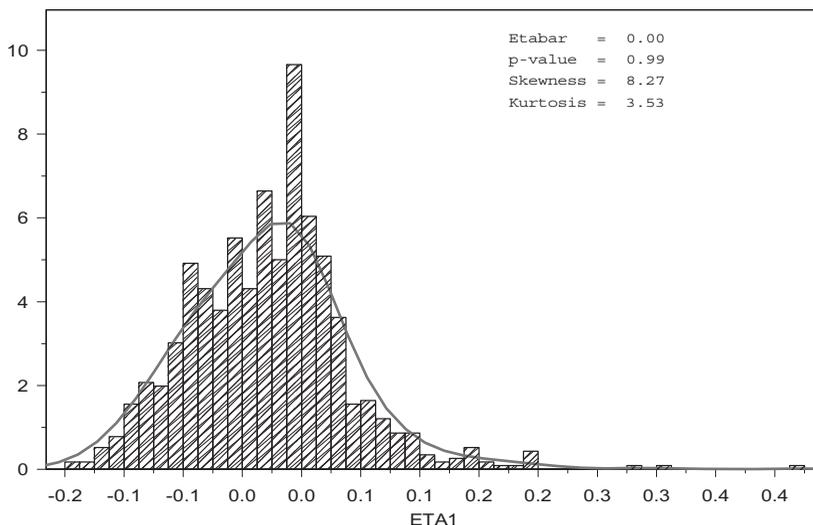
ENDIF
IF (MIXNUM.EQ.3) THEN
  AS=THETA(7)*EXP(ETA(3)) ; ; controls asymptotic weight gain
  W=THETA(3) ; residual error as a standard deviation
  K=THETA(5)*EXP(ETA(4)) ; controls rate of weight gain

ENDIF
  FA=BSLN*EXP(AS*(1-EXP(-K*TIME))) ; individual prediction
  Y=FA+W*EPS(1) ; additive residual error model

$MIX
  NSPOP=3 ; three subpopulations
  A=EXP(THETA(8))
  B=EXP(THETA(9))
  DEN=1+A+B
  P(2)=A/DEN ;PROB OF WEIGHT LOSS
  P(3)=B/DEN ;PROB OF WEIGHT GAIN
  P(1)=1/DEN ;PROB OF STABLE WEIGHT
```

The bounds on THETA(6) and THETA(7) ensure that the second and third sub-populations are associated with weight loss and weight gain, respectively. Additionally, interindividual variability is included on the parameters associated with the asymptotic weight gain or loss and the rate of weight gain. Despite dropping the only covariate (DOSE), the mixture model has resulted in a 5821 point drop in the objective function. The model now suggests that 20% of the patients are weight neutral, 32% are losing weight toward an asymptote of 96% of their baseline weight, and 48% are gaining weight toward an asymptote of 107% of their baseline weight. Now consider covariate inclusion in the submodels and probability model. The C9.TXT control stream produces a FIRSTONLY table where the covariates, modal random effect estimates, and EST (MIXEST) are summarized. First examine a box plot of baseline weight versus MIXEST and notice that those patients estimated to be losing weight tend to have higher baseline weights (Figure 28.5).

A similar box plot of DOSE versus estimated subpopulation suggests that those patients estimated to be gaining weight have much higher doses than do the other patients. Both of these findings suggest that baseline weight and DOSE might be included in our probability model. Turning attention to the submodels now, see that those patients estimated to be gaining weight have a monotone relationship between DOSE and ETA(3) (Figure 28.6). This suggests that DOSE be included as a covariate on the asymptote in the weight gain submodel.

FIGURE 28.5   A box plot of baseline weight by estimated subpopulation.

FIGURE 28.6   A box plot of random effect on weight gain asymptote versus dose.

Before the control stream is modified, the $CONTR record, which allows one to pass covariates to the $MIX block, is discussed. Typically, this record is positioned after the $INPUT record using the following syntax.

```
$CONTR DATA=(BSLN,DOSE)
```

This allows passage of the covariates baseline weight (BSLN) and a surrogate of exposure (DOSE) to the $MIX block. A word of caution is in order here. If the covariates being passed to $MIX appear after an alias assignment on the $INPUT record (i.e., DV=DSST or DATE=DROP), error messages may appear or the generated Fortran code for $MIX may be incorrect. If the data set has these properties, then a new one must be prepared. The following code modification illustrates the inclusion of the $CONTR record, modifications to the weight gain submodel to allow exposure to influence asymptotic weight gain, and modifications to the $MIX block to allow DOSE to influence the probability of being estimated as gaining weight and baseline weight to influence the probability of being estimated as losing weight (see C10.TXT R10.TXT).

```
$INPUT ID, DOSE, TIME, BSLN, AGE, SEX, SHOE, DV, EVID
$CONTR DATA=(DOSE,BSLN,SHOE) ; allows DOSE, BSLN, and SHOE to be
passed to $MIX
$PRED
  IF (MIXNUM.EQ.3) THEN
  AS=THETA(7)*EXP(DOSE*THETA(10)+ETA(3)) ; DOSE now influences
  ;asymptote for gainers
  W=THETA(3)
  K=THETA(5)*EXP(ETA(4))
  ENDIF
$MIX
  NSPOP=3
  E=THETA(12)*BSLN  ;  BSLN  now  influence  probability  of  losing
weight
  A=EXP(THETA(8)+E)
  B=EXP(THETA(9)+THETA(11)*DOSE) ; DOSE now influences probability
of gaining weight
  DEN=1+A+B
  P(2)=A/DEN
  P(3)=B/DEN
  P(1)=1/DEN
```

Inclusion of the new covariates results in a 631 point drop in the objective function value. The new model describes three subpopulations whose partition depends on DOSE and baseline weight (BSLN). The second subpopulation consists of patients who lose weight approaching an asymptotic weight of 95% of their baseline. The third subpopulation consists of patients who gain weight approaching an asymptote of 103% or 115% of their baseline weight for placebo or 4 mg/day, respectively. The partition between the weight neutral/weight loss/weight gain subpopulations shifts from 33%/34%/33% for placebo to 12%/13%/74% at a DOSE of 4 mg/day, given a baseline weight of 78 kg. The proportional weight changes from baseline versus time for the three subpopulations and the partition between the subpopulations as a function of exposure are shown in Figures 28.7 and 28.8. If one selects only those patients estimated to be weight gainers and calculates the average of the modal estimates for ETA(3), one arrives at a value of −0.0009, in contrast to that output by NONMEM (−1.1). Similar calculations for the other two random effects will illuminate discrepancies as well. It is left as an exercise for the reader

FIGURE 28.7 A plot of proportional weight change versus time and dose.

FIGURE 28.8 The probability of partitioning of subpopulations versus dose (mg/day).

to continue the model building to find a missing covariate, which will be revealed in Section 28.9.

28.7 MULTIPLE MIXTURES

Miller et al. (22) recently published the application of mixture modeling to count data. This novel analysis applied one probability partition across several pharma-

codynamic (PD) parameters while analyzing drug and placebo data in a "delinked" fashion. The "delinking" aspect of the model assumed separate models for the effect of drug or placebo on seizure count. As a result, the estimated drug effect included any placebo effect that might be present. A typical patient had a baseline seizure count and then up to three subsequent counts during drug or placebo treatment. The final model found one probability partition across the estimated parameters: baseline seizure count (B), maximum fractional reduction in seizure count (E_{max}), and placebo effect (PCB). The main idea presented in this section is that one probability partition applied to several parameters may not be the best one can do. For example, suppose the baseline is partitioned with 20% of the population having relatively high values and 80% having lower values, and 50% of the population has a relatively high E_{max} and the remainder has a lower E_{max}. Furthermore, suppose that one's E_{max} classification is independent of one's baseline classification. The idea here is that one probability partition cannot describe such a state of reality. What is needed are two separate partitions (hence the topic multiple mixtures) such that an individual can have any combination {low E_{max}, high E_{max}} × {low baseline, high baseline}. The data (DATA3.TXT) to be used consists of 1000 patients receiving doses (DOSE) of an imaginary drug, with equal proportions receiving 0, 1, 2, 3, or 4 mg/day. Each patient has count data collected as in the original study. The initial attempt to model this data, of course, will not use a mixture. Because the data are counts −2 times the log likelihood of the data (see C11.TXT AND R11.TXT) will be modeled as follows:

```
$PRED
  IF (NEWIND.NE.2) THEN ; if first record of a subject
    BSLN=DV ; get observed baseline for tabling
  ENDIF
  FLG = 0
  IF(TIME.GT.0.5)FLG=1 ; FLG is used to turn on placebo or drug
after baseline
  BASE=THETA(1)*EXP(ETA(1)) ; individual baseline estimate
  PLAC=THETA(2)  ; placebo effect
  EMAX=THETA(3)  ; maximum drug effect
  ED50=THETA(4)  ; DOSE for effect= EMAX/2
  D=DOSE*EMAX/(DOSE+ED50) ; EMAX model for drug effect
  CNT1=BASE*(1-FLG/(1+EXP(-(D+PLAC)))) ; compute expectation for
Poisson distribution
  ;Compute -2*ln(likelihood) using
  ;Stirlings formula for log DV factorial
  IF (DV.GT.0) THEN
    LDVFAC=(DV+.5)*LOG(DV)-DV+.5*LOG(6.283185)
  ELSE
    LDVFAC=0
  ENDIF
  B=LOG(CNT1)
  Y=-2*(-CNT1+DV*B-LDVFAC)
```

The first three lines of code allow one to extract and table the observed baseline seizure count for each patient. The indicator variable FLG turns on the drug or

placebo part of the model post baseline, and a logistic functional form is used to multiply the baseline seizure count to assure the positivity of the mean count and to allow the effect of drug to approach that of placebo in a continuous fashion. For this first attempt the mean baseline seizure count is 12.8 with placebo or 4 mg/day yielding a 10% or 70% reduction in seizure counts, respectively. It is left to the reader to examine the distributions of the observed baseline seizure frequency and the modal estimates of ETA(1). Hopefully one will observe right skewed or possibly bimodal distributions, which leads us to the idea of a mixture on some aspect of the model. Next, a mixture is applied to the baseline seizure counts. This approach (C12.TXT R12.TXT) results in a 102 point drop in the minimum value of the objective function, and the mixture now estimates that 88% of the patients have a baseline seizure count of approximately 11 with 12% having a baseline count of 52. One might indeed be tempted to be content with one probability partition and inject E_{max} into the mixture as follows (see C13.TXT R13.TXT):

```
IF (MIXNUM.EQ.1) THEN
  BASE=THETA(1)*EXP(ETA(1)) ; individual baseline count for first
                              subpopulation
  EMAX=THETA(3) ; EMAX for first subpopulation
    ELSE
  BASE=THETA(5)*EXP(ETA(2)) ; individual baseline count for second
                              subpopulation
  EMAX=THETA(7) ; EMAX for second subpopulation
ENDIF
```

Indeed, this seems to have been a savvy move as the minimum objective function value is now decreased by 3694 points. Now 29% of the population is estimated to have a baseline count of 13.5 and reduction in seizure frequency (including placebo effect) of 9% at a dose of 4 mg/day, and the remaining 71% of the population has an estimated baseline seizure frequency of 13.7 and a reduction of seizure frequency (including placebo effect) of 90% at a dose of 4 mg/day. The first mixture analysis (C12.TXT) suggested that there were two subpopulations with respect to baseline seizure count, and the second suggested that there were two subpopulations with respect to E_{max}. Hence, one might imagine that a patient could have any combination of low or high baseline and low or high E_{max}. This "multiple mixture" is implemented as follows (see C14.TXT R14.TXT):

```
IF (MIXNUM.EQ.1) THEN ; proportion of population with BASE=THETA(1)
                        and
  ;EMAX=THETA(3)
  BASE=THETA(1)*EXP(ETA(1))
  EMAX=THETA(3)
ENDIF
IF (MIXNUM.EQ.2) THEN ; proportion of population with BASE=THETA(5)
                        and
  ;EMAX=THETA(7)
  BASE=THETA(5)*EXP(ETA(2))
  EMAX=THETA(7)
ENDIF
```

```
IF (MIXNUM.EQ.3) THEN ; proportion of population with BASE=THETA(1)
                               and
  ;EMAX=THETA(7)
  BASE=THETA(1)*EXP(ETA(3))
  EMAX=THETA(7)
ENDIF
IF (MIXNUM.EQ.4) THEN ; proportion of population with BASE=THETA(5)
                               and
  ;EMAX=THETA(3)

  BASE=THETA(5)*EXP(ETA(4))
  EMAX=THETA(3)
ENDIF
```

This results in a drop in the minimum value of the objective function of 126 points. The results now indicate that 24%, 7%, 65%, or 5% of the population has estimated baseline seizure counts (percent reductions in seizure frequency) of 12 (10%), 54 (90%), 12 (90%), or 54 (10%), respectively. Assume that one allele pair {HH, hh, Hh, or hH} controls whether a patient has the high E_{max} {HH, Hh, or hH} in a dominant fashion (i.e., inheritance is dominant, monogenic, non-sex-linked). It is left as an exercise to calculate the gene frequencies for h and H (*answer*: frequency for H = 0.471 and 0.538 for h).

28.8   GRAPHIC CONSIDERATIONS

This section explores various diagnostic plots that might be used to support a mixture model. The issue of skewed random effects distributions as they relate to how well one can correctly classify patients into their correct subpopulations, assuming a mixed population, is also revisited.

A plot of observations versus predictions is often used as a measure of goodness of fit during model building. Returning to the first example, compare these plots for the nonmixture (C5.TXT) and mixture analysis (C4.TXT) in Figures 28.9 and 28.10, respectively.

In these figures the observed response is plotted on the abscissa and the prediction on the ordinate, so that one can easily see the distribution of predictions for a given level of response. In Figure 28.9 there is pronounced overprediction (underprediction) at lower (higher) observed values. Figure 28.10 leads one to believe that the fit is much improved relative to Figure 28.9. The predicted values output by NONMEM V when a mixture model is used are those PREDs for a patient's most likely subpopulation. This can be thought of as a type of "individualized" prediction, but using a marginal posterior estimate of the most likely subpopulation versus a modal or post hoc estimate of a random effect. In any case, such plots may be overly optimistic. One approach is to construct a plot of the expected predictions (or individualized predictions) versus observations. This expectation for a given prediction (*E(PRED)*) is defined as follows:

$$E(PRED) = \sum_{i=1}^{n} p(\theta)_i \cdot F(\theta, \eta)_i \tag{28.8}$$

FIGURE 28.9   Nonmixture model: a plot of predictions versus observations—all $\eta = 0$.

FIGURE 28.10   Mixture model: a plot of predictions versus observations—all $\eta = 0$.

Here, n is the number of subpopulations, $p(\theta)_i$ is the estimated probability that the patient belongs to the ith subpopulation, which may depend on fixed effects parameters, and $F(\theta, \eta)_i$ is the prediction for the ith subpopulation. For this computation one may evaluate $F(\theta, \eta)_i$ at $\eta = 0$ to get the expected prediction or at η equal to a post hoc or modal value to get the expected individualized prediction. NONMEM supports the calculation of these expectations and to communicate the need for these expectations requires one to modify the original control stream to include the $ABBREVIATED record ($ABB) and a block of code to compute the expectations. Modify C4.TXT to make C15.TXT, and the $ABB record looks like

```
$INPUT ID, AMT, TIME, WT, AGE, DV, EVID, MDV REA
$ABB COMRES=1 COMSAV=1
```

COMRES=1 instructs NONMEM to reserve the first position in NMPRD4 (a NONMEM PRED common) for variable storage and this position will be referenced as COM(1) in abbreviated code. COMSAV=1 instructs NONMEM to handle COM(1) differently depending on whether the reserved variable COMACT is 1 or 2. The block of code to compute the expectations is placed within the $ERROR block as follows:

```
$ERROR
  Y=F*(1+EPS(1))
IF (COMACT.EQ.1) THEN
  IF (MIXNUM.EQ.1) COM(1)=0
  COM(1)=COM(1)+MIXP(MIXNUM)*Y
ENDIF
```

The test for COMACT.EQ.1 instructs NONMEM to use final THETA estimates and set all η random effects to zero. The code then initializes the expectation to zero for each new prediction and computes the value defined by Eq. (28.8) and stores it in COM(1), which will then be tabled for plotting (T15.TXT). Inspection of T15. TXT reveals very low expectations for the first patient. As it turns out this is a bug that involves the use of METHOD=1 or the POSTHOC option with METHOD=0. The work around for the METHOD=0 is to remove the POSTHOC option, and for METHOD=1 is to change to METHOD=0 and MAXEVALS=0 and use the final estimation results from the METHOD=1 analysis. When the POSTHOC option (C16.TXT R16.TXT) is removed, the expected predictions for the first patient have changed substantially (T16.TXT) and these are shown in Figure 28.11.

FIGURE 28.11 Mixture model: a plot of expected predictions versus observations—all $\eta = 0$.

The visual goodness of fit for Figure 28.11 appears closer to that seen in Figure 28.9 than seen in Figure 28.10. However, when one calculates the mean error (observed minus predicted or observed minus expected prediction), one finds less bias when the expected prediction is used.

To calculate expected individualized predictions, one needs to test for COMACT. EQ.2 in the above code. This has the effect of using final thetas and conditional (post hoc or modal) etas in the computation. When this code is used with METHOD=0, one must include the POSTHOC option on the $ESTIMATION record (see C17.TXT R17. TXT).

One can use the *central limit theorem* to compare goodness of fit between a nonmixture model and a mixture model. If there is a covariate that can be used to group the observations such that there are a substantial number of observations per group, and the expectations of each member of the group are equal, then the average of the observations within a group should be approximately normally distributed about the expected observation for the group. The goal is to plot averaged predictions (for the nonmixture model) or averaged expected predictions (for the mixture model) versus the averaged observations for each group.

Choose time as the grouping covariate. The original data set DATA1.TXT needs some modification to take advantage of NONMEMs RAW_ data item, which is used for calculating the group averages. Specifically, transform DATA1.TXT to DATA4.TXT by adding a column for the RAW_ data item, adding a "dummy" individual at the start of the data set to serve as a template to tell NONMEM how to calculate the group averages, and then reorder the columns so that the grouping will be done based on time. The RAW_ data item is placed in the rightmost column of the data set, and it has two allowable values. If a zero is present for RAW_, then tabled or plotted values for the DV associated with that record will be those present in the original data set. This is the case for the records of the original 100 subjects. If a one is present in RAW_, then MDV for that record must also equal one (although EVID can equal zero), and this indicates that this record serves as a template record for our grouping and averaging process. Inspection of our data set shows that RAW_=1 and MDV=1 for all but the first record (the dosing record) for our dummy individual (ID=0). The columns have been reordered such that the only non-NONMEM data item to the left of DV on the $INPUT record is TIME. This has the effect of setting the DV items for all but the first row of the dummy subject to the average of the DV values for the records that match the template row. For example, the tabled or plotted observed value for the second row of the dummy subject will be the average value of the observations with TIME values equal to one (using only those records with MDV=0). For this example the grouping is only based on time. Other non-NONMEM defined items could be moved to the left of the DV column to define different groups, although for this example none of them make much sense as all the doses are the same, whereas age and weight assume too many values to be useful. NONMEM data items (MDV and RAW) are not used for matching observation records with the template records, and it does not matter which side of the DV column they reside on. Any record that matches in the user defined data items to the left of the DV column will be used for the averaging. If there are multiple records per subject that match a template record, then NONMEM calculates a two-stage average, averaging first within the subject and then across subjects. PRED defined data items that are stored in the SAVE region are averaged as are the DV items, provided that mixture modeling is not being

attempted. These data and prediction averaging processes could certainly be done outside NONMEM, but the example will show how to handle both cases (mixture and nonmixture) using NONMEM.

Proceed to modify the nonmixture model control stream (C5.TXT) resulting in C18.TXT. Important aspects of the new control stream are the modifications to $INPUT to assure that grouping is by TIME.

```
$INPUT ID, TIME, DV, AMT, WT, AGE, EVID, MDV, REA, RAW_
$ABB COMRES=1 COMSAV=1
$DATA DATA4.TXT (10E12.0) ;NOOPEN REWIND
$DATA IGNORE=#
```

In the $ERROR block calculate the expected prediction as follows:

```
$ERROR
  COM(1)=F
  Y=F*(1+EPS(1))
```

Recalling that the presence of the POSTHOC option on the $ESTIMATION record causes incorrect expected predictions for the first subject (in this case the most important subject) with METHOD=0 in the presence of a mixture model, one might be tempted to remove it. For nonmixture models, where the expected prediction is the prediction, the presence of the POSTHOC option does not seem to influence the output. The changes to the data structure and the code instructing NONMEM to produce a scatterplot containing only points for the first individual, and an output table from which we can make our own version of the scatterplot, are summarized below.

```
$ESTIMATION MAXEVALS=9000 PRINT=1 ; METHOD=1 INTERACTION
$COV
$SCAT DV VS COM(1) UNIT
$TABLE ID DV PRED TIME MDV COM(1) ONEHEADER NOPRINT FILE=T18.TXT
```

The scatterplot produced by NONMEM will display 24 points. Each point will correspond to one time point and have as its coordinates the average of the observations and the average of the predictions for that time point. To create a separate scatterplot, one would just plot DV versus COM(1) for the first (dummy) patient. In this example we have chosen to set COM(1)=F. We also might have set COM(1)=Y. As it turns out this latter approach produces incorrect output. Here, because we are using METHOD=0, F is equal to PRED. If one were using some type of conditional estimation, F would be evaluated at modal values of η and would not equal PRED. One way to trick NONMEM into producing the desired output would be to use the final model estimates from conditional estimation as initial values for a METHOD=0 run with MAXEVALS=0.

For the mixture model a bit more work is needed to get the job done. Because averaging PRED defined data items is problematic when a mixture is employed, one might be tempted to use two $PROBLEMs concatenated. The first might generate the expected predictions under the mixture model and table them, and the second could

read in this table and do the averaging with a nonmixture model. While this seems elegant, it is generally desirable to restrict all abbreviated code to the first $PROBLEM. Rather than explore this, a slightly longer but simpler path is taken. First, calculate the expected predictions and write them out into the last column of a table using the data structure of DATA4.TXT. This only involves the following new code (see C19.TXT which is derived from C4.TXT):

```
$TABLE ID, TIME, DV, AMT, WT, AGE, EVID, MDV, REA, RAW_
  COM(1) NOAPPEND NOHEADER NOPRINT FILE=T19.TXT
```

Next, read in T19.TXT with a new nonmixture control stream (C20.TXT) and produce a table containing the averaged expected predictions from the mixture analysis (C19.TXT). Important features of C20.TXT include modifications to accommodate the new data file, which has an extra column at the end for the expected predictions, named CM.

```
$INPUT ID, TIME, DV, AMT, WT, AGE, EVID, MDV, REA, RAW_ CM
$ABB COMRES=1 COMSAV=1
$DATA T19.TXT (11E12.0) ;NOOPEN REWIND
```

Then place CM in the SAVE region so it will be averaged like the DV.

```
$ERROR
  Y=F*(1+EPS(1))
  COM(1)=CM
```

Finally make a table from which DV versus COM(1) will be plotted for the first subject.

```
$TABLE ID, TIME, DV, AMT, WT, AGE, EVID, MDV, REA, RAW_
  COM(1) ONEHEADER NOPRINT FILE=T20.TXT
```

Taking the output from these runs (C18.TXT and C20.TXT) and plotting the averaged observations versus averaged predictions or averaged expected predictions for the nonzero time points for the dummy individual gives Figures 28.12 and 28.13, respectively.

Here, bias is seen in the plot from the nonmixture model, supporting the notion that the mixture model might be superior.

For the last topic in this section, the situation where we have a skewed modal eta distribution is revisited, but this time there is no model misspecification. Using techniques from Section 28.9, simulate and then estimate rich PK data as in the first example (100 subjects, 2400 observations), such that 80% of the population has a K_a of 1 and the remaining 20% of the population has a K_a of 4. Both subpopulations have interindividual variability in K_a consistent with a CV of 47%. Using a two $PROBLEM control stream (C21.TXT) and data skeleton (DATA5.TXT), first simulate the data and record as part of the data which subpopulation each patient is assigned to with the following code:

FIGURE 28.12   A plot of expected prediction versus observation averaged by time (non-mixture model).

FIGURE 28.13   A plot of expected prediction versus observation averaged by time (mixture model).

FIGURE 28.14   Density of modal values for random effect on K_a (for subpopulation with high K_a) when simulation and estimation models are identical.

```
IF (ICALL.EQ.4) THEN
  REA=MIXNUM
ENDIF
```

Then during estimation record which subpopulation is most likely for each patient using

```
EST=MIXEST
```

This allows one to keep track of patients and see how they might be misclassified during estimation (see `T21A.TXT`). With this example, even though there is no model misspecification, one sees that many patients simulated to have the lower value of K_a are estimated to have the higher of the two values. These patients who ideally would be estimated to be in the subpopulation associated with `ETA(2)` instead tend to have slightly negative values of `ETA(3)` as they have been incorrectly assigned to the second subpopulation. The net effect here is that the density of `ETA(3)` is inappropriately augmented for negative values, giving the impression of skewness (see Figure 28.14).

This undesirable result can likely be attributed to the magnitudes of inter- and intraindividual variability used during the simulation step.

28.9   SIMULATION WITH MIXTURES

The focus of this section is on some of the code that has been used to generate the data that has been used for the previous examples and on exploring how one might automate a posterior predictive check with or without a bootstrap step for the weight change example.

The control stream/skeleton data set pairs used to simulate the first PK example, the weight change example, and the multiple mixture seizure count example can be found in the book's ftp site in the following files: `C22.TXT`/`SKEWDATA.TXT`, `C23.TXT`/`MIXDATA1.TXT`, and `C24.TXT`/`MIXDATA2.TXT`, respectively. For each of these control stream/data set skeleton pairs the control stream simulates a new data set with identical structure to the skeleton, but with the DV simulated based on the fixed and random effects parameters in the control stream.

For the closing exercise we explore how to automate a type of model evaluation tool called posterior predictive check (PPC). This technique has been developed and described elsewhere (32–34). The PPC technique has been used as both a tool to monitor model development and to validate (i.e., determine the predictive performance of) final models. The basic idea is to construct a nonsufficient statistic that is of clinical or scientific interest and that would not automatically fit the model under investigation, and compute this statistic for the observed data set and for many data sets that have been simulated using the model (32). Then the distribution of statistics from the simulated data sets is compared to the statistic computed from the observed data. This comparison might be as simple as plotting the statistics from the simulated data sets as a histogram and then plotting the observed statistic as a vertical line. On the other hand, a more formal approach could be employed whereby a *p*-value is calculated as the probability that the simulated statistic is more extreme than the value of the statistic calculated from the original data (34). This latter approach will not be attempted here.

Using the weight change example, define the nonsufficient statistic to be the maximum weight gain (`DV-BSLN`) observed in a simulated or the original data set. The PPC procedure can be undertaken several ways. The simplest way would be to compute the statistic for the original data set and then repeatedly simulate data sets using the data structure for the original data set and compute the statistic for each of these simulated data sets. This simple approach does not take into account the uncertainty in the parameter estimates from our model (i.e., information from the `$COVARIANCE` step is not utilized). Our approach takes this uncertainty in parameter estimates into account, and it is sometimes referred to as a PPC with bootstrap step (33). The solution to this problem will allow the reader to review several nice features of the NONMEM program such as the notion of a `$SUPER` problem, model specification input and output records (`$MSFI` and the `MSFO` option for the `$ESTIMATION` record), verbatim code, and the `PASS` utility.

All these aspects of the problem are discussed as they arise, but first an overview of what we wish to accomplish is provided.

1. Starting with a dummy `$PROBLEM`, our test statistic for the original data set is calculated and written to a file. This problem includes a `$SIMULATION` record so that the subsequent problem can use a special random seed so that each of the simulated data sets will be different.
2. The second `$PROBLEM` simulates a data set based on the structure of the original data set and then performs estimation on this simulated data set and saves the results of this estimation in a model specification output file (`MSFO`).
3. The third `$PROBLEM` simulates a final data set based on the parameter estimates in the `MSFO` from the previous problem.

4. The final $PROBLEM calculates the statistic for the data set simulated in step 3 and appends it to the file created in step 1.

5. Steps 2, 3, and 4 are repeated 30 times.

The result is a file with the first row containing that statistic for the original data and the next 30 rows containing the statistics from the simulated data sets. In practice, one would usually construct more than 30 simulated data sets.

The first $PROBLEM in the "mega control stream" (C25.TXT) contains all of the abbreviated and verbatim code. The abbreviated code is just that from the original model (C10.TXT). Modifying this control stream by adding some verbatim code allows one to output the statistics to a file called tj.

```
" FIRST
" INTEGER unitp
" CHARACTER cisym * 40
" unitp=42
" cisym='tj'
```

Next, insert a block of code that calculates the statistic for the original data set. When multiple $PROBLEMs are present in a control stream, they are indexed with the variable IPROB. This block of code is only invoked for the first $PROBLEM at its initialization (ICALL=1). The PASS utility is used to read through DATA2.TXT and find the largest value for (DV-BSLN), store it in the variable AV, and then write it to the file. In the current example, all rows of DATA2.TXT contain observations so the code is fairly simple. The code could be modified to compute several statistics, perhaps grouping by dose.

```
IF (IPROB.EQ.1.AND.ICALL.EQ.1) THEN
  AV=-1
  MODE=0
  CALL PASS (MODE)
  MODE=2
  CALL PASS (MODE)
  DOWHILE(MODE.NE.0)
    IF ((DV-BSLN).GT.AV) THEN
      AV=DV-BSLN
    ENDIF
    CALL PASS (MODE)
  ENDDO
" OPEN (42,FILE=cisym, ACCESS='append')
" call files(unitp)
" write(42,100) AV
" CLOSE(unitp)
" call files (unitp)
" 100 format(F12.4)
ENDIF
```

A second block of code almost identical to the above code is used to calculate the same statistic for the last (IPROB=4) $PROBLEM at its finalization (ICALL=3). The

reader is encouraged to read *NONMEM User's Guide 8* for an in-depth discussion of the PASS utility and verbatim code. In the above code a Fortran "write" statement is used within the verbatim code to write the data to the file cisym. The only other noteworthy aspect of the first $PROBLEM is the following:

```
$SIM (93) ONLYSIMULATION
```

Its only purpose is to allow one to have a $SIM record in the subsequent $PROBLEM with a random seed equal to –1.

The next item of interest is the "super problem" record.

```
$SUPER SCOPE=3 ITERATIONS=30
```

This instructs NONMEM to expect three $PROBLEMs to follow and to run all three of them in their order of occurrence 30 times. The first of these uses DATA2.TXT as the infile, which must be rewound.

```
$DATA DATA2.TXT REWIND
```

In this $PROBLEM, data is first simulated with a random seed continued from the previous $PROBLEM and then estimated, with the estimation results saved in a model specification output file as follows:

```
$SIM (-1)
$ESTIM MAXEVAL=9000 PRINT=1 NOABORT METHOD=1 MSFO=M25
```

For the next $PROBLEM, simulate and $TABLE the final data set, which is used to calculate the test statistic. Relevant code is as follows:

```
$SIMULATION (-1) ;ONLYSIMULATION
$TABLE ID DOSE TIME BSLN AGE SEX SHOE DV EVID
  NOPRINT NOHEADER NOAPPEND FILE=SIMU25.TBL
```

It is important to note that the data set that we simulate has the same order of appearance of variables as the original data set. This is accomplished by copying the variable names from the $INPUT record and then using the NOAPPEND option so that DV, PRED, RES, and WRES are not appended. When working with a large data set with many columns that are not being used, one might try to simulate a more "compact" data file, containing only the columns needed for the PPC statistic calculation. Caution must be used here. Suppose that a table is simulated as above, but that the original data set had the following $INPUT record:

```
$INPUT ID DOSE DV TIME ALB=DROP BSLN AGE SEX SHOE EVID
```

The code used to calculate the test statistic on the original data would stay the same but the code used to calculate the statistic for the simulated data set would need to be changed as follows. BSLN and DV are the fourth and eight variables in the simulated data set. Ignoring DROPed variables TIME and SHOE are the fourth and eight

FIGURE 28.15 Posterior predictive check with bootstrap step for weight change mixture problem.

variables in the original data set. Therefore, one would substitute TIME for BSLN and SHOE for DV in this block of code.

Finally, for the fourth $PROBLEM the new simulated data is read in, after it is rewound and the statistic is calculated.

```
$DATA SIMU25.TBL (9E12.0) NOOPEN REWIND
```

Taking the output file "tj", a histogram is plotted, using the last 30 rows (which reflect simulation), and the data from the first row is added as a vertical line to reflect the observed statistic (Figure 28.15).

When using the PPC with bootstrap step for model monitoring, it may be difficult to decide whether to include the bootstrap step, which involves estimating a simulated data set, which may be quite time consuming. Indeed, for models that can be written out explicitly (i.e., differential or algebraic equation solvers are not needed), one might be better off using standard statistical software to simulate the data and compute the statistics. Doing this, information from the $COV step could be used, without having to perform the time-consuming data estimation step within NONMEM. Some older versions of NONMEM V will generate the following Fortran error with this control stream.

```
Severe (620): Too many bytes read from unformatted record.
```

If this occurs, one needs to contact the Globomax Icon LLC (http://www.globomaxservice.com/) and request a replacement DAT1.FOR routine.

28.10   SUMMARY

Several examples of mixture modeling using NONMEM V are explored. The need to use mixture models generally arises when there is some missing covariate in the data set or the distribution of a random effect is better modeled as a mixture of symmetric distributions than with a single distribution. Advances in molecular biology and pharmacogenomics may someday yield data sets where mixtures are not needed. Mixture modeling should generally be used only after considerable work has been done to optimize structural, random effects, and covariate aspects of a nonmixture model, although some authors advocate introducing the mixture prior to covariate selection.

REFERENCES

1. E. Peretti, G. Karlaganis, and G. H. Lauterburg, Acetylation of acetylhydrazine, the toxic metabolite of isoniazid in humans. Inhibition by concomitant administration of isoniazid. *J Pharmacol Exp Ther* **243**:686 (1987).

2. W. W. Weber and D. W. Hein, N-acetylation pharmacogenetics. *Pharmacol Rev* **37**:25 (1985).

3. F. J. Chaves, V. Giner, D. Corella, J. Pascual, P. Marin, M. E. Armengod, and J. Redon, Body weight changes and the A-6G polymorphism of the angiotensinogen gene. *Int J Obes Relat Metab Disord* **26**:1173–1178 (2002).

4. G. P. Reynolds, Z. J. Zhang, and X. B. Zhang, Association of antipsychotic drug induced weight gain with a 5-HT2C receptor gene polymorphism. *Lancet* **359**:2086–2087 (2002).

5. H. Y. Lane, Y. C. Chang, Y. C. Cheng, G. C. Liu, X. R. Lin, and W. H. Chang, Effects of patient demographics, risperidone dosage, and clinical outcome on body weight in acutely exacerbated schizophrenia. *J Clin Psychiatry* **64**:316–320 (2003).

6. M. Masellis, V. Basile, H. Y. Meltzer, L. A. Lieberman, S. Sevy, F. M. Macciardi, P. Cola, A. Howard, F. Badri, M. M. Nothen, W. Kalow, and J. L. Kennedy, Serotonin subtype 2 receptor genes and clinical response to clozapine in schizophrenia patients. *Neuropsychopharmacology* **19**:123–132 (1998).

7. J. Marc, J. Prezelj, R. Komel, and A. Kocijancic, VDR genotype and response to etidronate therapy in late postmenopausal women. *Osteoporos Int* **10**:303–306 (1999).

8. J. A. Millar, K. Thai, and J. W. Scholey, Angiotensin II type 1 receptor gene polymorphism predicts response to losartan and angiotensin II. *Kidney Int* **56**:2173–2180 (1999).

9. P. Dalen, M. L. Dahl, M. L. B. Ruiz, J. Nordin, and L. Bertilsson, 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. *Clin Pharmacol Ther* **63**:444–452 (1998).

10. F. Saarikoski, K. Sata, J. Husgafvel-Pursianen, O. Haukka, J. Impivaara, H. Jarvisalo, H. Vainio, and A. Hirvonen, CYP2D6 ultrarapid metabolizer genotype as a potential modifier of smoking behavior. *Pharmacogenetics* **10**:5–10 (2000)

11. P. Dalen, C. Frengall, M. L. Dahl, and F. Sjoqvist, Quick onset of severe abdominal pain after codeine in an ultrarapid metabolizer of debrisoquine. *Ther Drug Monit* **19**:543–544 (1997).

12. J. A. Kuivenhoven, J. W. Jukema, A. H. Zwinderman, P. Knijff, R. McPherson, A. V. G. Bruschke, K. I. Lie, and J. J. P. Kastelein, The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. *N Engl J Med* **338**:86–93 (1998).

13. D. K. Kim, S. W. Lim, S. Lee, S. E. Sohn, S. Kim, C. G. Hahn, and B. J. Carroll, Serotonin transporter gene polymorphism and antidepressant response. *Neuroreport* **11**:215–219 (2000).

14. A. Odani, Y. Hashimoto, Y. Otsuki, Y. Uwai, H. Hattori, K. Furusho, and K. Inui, Genetic polymorphisms of the CYP2C sub-family and its effect on the pharmacokinetics of phenytoin in Japanese patients with epilepsy. *Clin Pharmacol Ther* **62**:287–292. (1997).

15. K. Mamiya, I. Ieiri, J. Shimamoto, E. Yukawa, J. Imai, H. Ninomiya, H. Yamada, K. Otsubo, S. Higuchi, and N. Tashiro, The effect of genetic polymorphisms of CYP2C9 and CYP2C19 on phenytoin metabolism in Japanese adult patients with epilepsy: studies in stereoselective hydroxylation and population pharmacokinetics. *Epilepsia* **39**:1317–1323 (1998).

16. S. L. Beal and L. B. Sheiner, *NONMEM Users Guides*. NONMEM Project Group, University of California, San Francisco, 1999.

17. K. Pearson, Contributions to the mathematical theory of evolution. *Phil Trans R Soc Lond A* **185**:71–110 (1894).

18. J. P. Harding, The use of probability paper for the graphical analysis of polymodal frequency distributions. *J Marine Biol Assoc UK* **28**:141–153 (1948).

19. C. R. Rao, The utilization of multiple measurements in problems of biological classification. *J R Stat Soc B* **10**:159–203 (1948).

20. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

21. B. Facca, B. Frame, and S. Triesenberg, Population pharmacokinetics of ceftizoxime administered by continuous infusion in clinically ill patients. *Antimicrob Agents Chemother* **42**:1783–1787 (1998).

22. R. Miller, B. Frame, B. W. Corrigan, P. Burger, H. Bockbrader, H. Garofolo, and R. Lalonde, Exposure–response analysis of pregabalin add-on treatment of patients with refractory partial seizures. *Clin Pharmacol Ther* **73**:491–505. (2003).

23. R. Hussein, B. G. Charles, R. G. Morris, and R. L. Rasiah, Population pharmacokinetics of perhexiline from very sparse, routine monitoring data. *Ther Drug Monit* **23**:636–643 (2001).

24. T. Shiiki, Y. Hashimoto, and K. Inui, Simulation for population pharmacodynamic analysis of dose-ranging trials: usefulness of the mixture model analysis for detecting nonresponders. *Pharm Res* **19**:909–913 (2002).

25. P. H. Zingmark, M. Ekblom, T. Odergren, T. Ashwood, P. Lyden, M. Karlsson, and N. Jonsson, Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol* **56**:173–183 (2003).

26. R. D. Gibbons, E. Dorus, D. G. Ostrow, G. N. Pandey, J. M. Davis, and D. L. Levy, Mixture distributions in psychiatric research. *Biol Psychiatry* **19**:935–961 (1984).

27. G. J. McLachlan and K. E. Basford, *Mixture Models—Inference and Applications to Clustering.* Marcel Dekker, New York, 1988.

28. D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. Wiley, Hoboken, NJ, 1985.

29. B. Frame, R. Miller, and R. Lalonde, Evaluation of mixture modeling with count data using NONMEM. *J Pharmacokinet Pharmacodyn* **30**:167–183 (2003).

30. R. I. Jennrich, Asymptotic properties of non-linear least squares estimators. *Ann Math Stat* **40**:633–643 (1969).

31. J. H. Wolfe, A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinomial distributions. *Technical Bulletin STB 72-2*. U.S. Naval Personnel and Training Research Laboratory, San Diego, 1971.

32. T. R. Belin and D. B. Rubin, The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Stat Med* **14**:747–768 (1995).

33. P. O. Gisleskog, M. O. Karlsson, and S. L. Beal, Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn* **29**:473–505 (2002).

34. Y. Yano, S. L. Beal, and L. B. Sheiner, Evaluating pharmacokinetic/pharmacodynamic models using the posterior predicitive check. *J Pharmacokinet Pharmacodyn* **28**:171–192 (2001).

CLINICAL TRIAL DESIGNS

Designs for First-Time-in-Human Studies in Nononcology Indications

HUI-MAY CHU, JIUHONG ZHA, AMIT ROY, and ENE I. ETTE

29.1 INTRODUCTION

The primary objective of first-time-in-human (first-time-in-man or first-in-human) studies is the identification of a suitable dose or dose range for further study, based on the safety and tolerability of the substance. These are volunteer subject-based dose-escalation studies that are traditionally small and time-lagged. These studies offer the first opportunity to learn about the drug in humans and serve as a bridge from animal to human. They provide opportunity for confirming the prediction of pharmacokinetics from animals and to learn about the safety of the drug, if the study is appropriately designed.

The *Critical Path Initiative* proposed by the Food and Drug Administration (1) has emphasized the need for informative knowledge-based drug development, and there has been much focus on the optimization of early drug development (2, 3). Phase 1 studies are critical in the overall process of drug development, and the need for the implementation of informative designs has been advocated in the literature (4–8).

It is challenging to get an overview of studies performed since there is little written on the design of first-time-in-human studies, and no consensus exists on how the studies should be designed. Buoen et al. (9) did a literature survey study of 105 studies comprising 3323 healthy volunteers published in the five major clinical pharmacology journals between 1995 and 2004 and found that the average trial was a placebo-controlled, double-blind study. Such a trial had five dose levels with a sample size of 32 subjects, but there was great variation in cohort size and dose-escalation method used in the studies. The most common design was the parallel single-dose design. The crossover design, on the other hand, was more frequent in early publications of the period of survey reported.

The literature is replete with scientific study designs for Phase 1 cancer trials in patients (e.g., see Ref. 10), but such is not the case for first-time-in-human (FTIH)

studies for nononcology indications. To be able to understand how to design Phase 1 nononcology studies, it is important to understand how they are currently designed. Currently, there are as many study designs for FTIH nononcology studies as there are pharmaceutical companies, and the design chosen is more a result of habit and preference than a well-founded scientific and pharmacometric rationale. Not all FTIH studies are published (11), and there is very scanty information on how to design these studies efficiently. Thus, it is challenging to obtain an overall picture of how these studies are designed and how well various designs perform.

29.2   DOSE-ESCALATION SCHEME

Buoen et al. (9) reported that the dose-escalation schemes used in FTIH studies could be categorized as linear, logarithmic, modified Fibonacci, or miscellaneous. The latter included dose-escalation regimens in which the three standardized methods are combined. The authors reported that in 12 out of the 105 studies they reviewed a linear escalation method with fixed dose increment was used. A logarithmic dose-escalation scheme in which the relative dose increment was the same (e.g., 100%) was used in 22 studies. Four of the studies used a modified version of the Fibonacci escalation scheme, which is frequently used in cancer Phase 1 trials (6, 12–14). For most of the studies reviewed (i.e., 63.8%, or 67 studies) the dose-escalation schemes used did not seem to follow one particular scheme. In some cases two of the escalation schemes described above were combined (e.g., starting with a logarithmic escalation to convert later into a modified Fibonacci sequence), while for other studies, no escalation scheme was apparent. The doses appeared to have been chosen arbitrarily (11).

29.3   FIRST-TIME-IN-HUMAN STUDY DESIGNS

Five major types of study design have been identified as those commonly used in the FTIH studies (11). The parallel dose design in which each subject enrolled in a study receives only one administered dose of the drug was described as being the most common. Escalation to a new dose with new cohort of subjects is accomplished after a safety evaluation. Fewer studies reported in the literature use either the fixed sequence escalation design in which several different doses were administered to each subject, or the alternating crossover dose-escalation designs (11). The crossover designs are more efficient for estimating dose proportionality, and they allow for more information to be obtained from fewer subjects (15, 16). In evaluating the efficiency of FTIH design alternatives in estimating dose proportionality using mixed effects modeling, it was concluded that an alternate crossover design always required fewer subjects to achieve the same precision when compared with a sequential design (17).

29.4   COHORT SIZE IN FTIH STUDIES

Habit and preference form the basis for choosing cohort sizes for FTIH healthy volunteer studies. In reviewing studies reported in the literature, it was observed

that the rationale for the choice of cohort sizes is not usually provided (18). Also, it has been reported that the probability of observing spontaneous adverse events such as elevation of markers of hepatotoxicity aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (AP), or gamma-glutamyl transpeptidase (γGT) to two times the upper limit of the normal range (ULN) is 6.71% (18). Elevation of these enzymes is indicative of hepatocellular injury (19), and their presence in serum is considered to be reliably indicative of a recent hepatocyte injury (20). An increase in one of these enzyme levels to $2 \times$ ULN is defined by international consensus as an indication of clinically relevant hepatic injury (21). It has been shown that for large cohort sizes, the probability of an event occurring in an actively treated subject is so large that the probability of a Type I error exceeds the preferred level of significance (0.05) (18). However, in small cohorts there is much to gain with the inclusion of one extra subject when the cohort size is less than six active subjects. This is because with less than six active subjects in a cohort, the active events that are detectable with a given power decrease quickly with increasing cohort sizes. The authors inferred that the cohort size in Phase 1 studies should not be less than six active subjects (18). They noted that the gain is smaller for larger cohort sizes. Little is to be gained by increasing the number of subjects if the cohort has more than ten active subjects.

29.5 ADVERSE EVENTS IN FTIH STUDIES

Sibille et al. (22) reported on clinical, laboratory, and electrocardiographical adverse events detected in healthy volunteers in a Phase 1 center over a 10 year period. The data base covered 54 Phase 1 studies that involved 1015 healthy young volunteers (993 males) who received 1538 treatments (23 different active drugs or placebo) corresponding to 12,143 days of follow-up. They observed that the overall incidence of adverse events was 12.8%, and a significant difference between active-drug (13.7%) and placebo (7.9%) treatments was obtained. One thousand five hundred and fifty eight adverse events of 110 distinct types were documented. The incidence was superior to 10% in only three adverse events (headache, diarrhea, and dyspepsia). Subjects on placebo also experienced most of these adverse events. Three percent of the adverse events were rated as severe, while 97% were of minor intensity. Some of the adverse events were related to a vagal reaction or to study conditions, but not the tested drugs. There was no life-threatening event or death. The global rate of adverse event occurrence was one per treatment. For two successive 5 year periods, no difference in the overall incidence of adverse events with placebo was observed. The authors concluded that adverse events in Phase 1 studies are very common. However, these were usually of minor intensity and rarely severe; even though exceptional, life-threatening adverse events are possible (22).

29.6 DETERMINATION OF THE EFFICIENCY OF
FTIH DESIGNS IN HEALTHY VOLUNTEERS

FTIH studies present the drug developer/clinical researcher with the first opportunity to study a new molecular entity (NME) in humans and are pivotal to early

clinical development of the NME. If designed appropriately, a lot of knowledge can be gained from such studies.

There has been an increased focus on augmenting the productivity of the pharmaceutical industry through the application of knowledge-based approaches to the drug development process without compromising safety. These efforts have resulted in "The Critical Path" in the United States, and the "New Safe Medicines Faster" in Europe (23). Phase 1 FTIH studies provide an excellent opportunity for learning about the NME's safety and tolerability in humans, confirming predicted pharmacokinetics from animals and learning more about the pharmacokinetics in humans.

Not much has been written on the design of FTIH studies for nononcology indications, but there is need for rapid and efficient dose-escalation schemes based on both pharmacodynamics (safety and biomarkers if present) and pharmacokinetics. In view of the small numbers of healthy volunteers enrolled in early phase development clinical trials, it is important to choose an efficient experimental design that takes in the safety and pharmacokinetic (PK) objectives (e.g., dose proportionality) of the trial. In general, crossover designs have been recommended (17). However, the analysis of data from a crossover design poses several problems, including nonconstant variances for all observations and the possibility of carryover effects (16). A major factor to be considered in the design of a FTIH study is safety. Dose escalation cannot proceed unless safety at a current lower dose is established. It is also important in a FTIH study design to be able to address the question of dose proportionality and precision in PK parameter estimates. FTIH studies must be conducted within a reasonable time for mission-critical decision making about the advancement of the NME in development. Hence, study duration and study size are also important. Thus, there can be a plethora of FTIH study designs that can be and have been developed to address these issues. It is therefore important in choosing between designs to take into account the efficiency of these designs in addressing these issues—safety, dose proportionality, precision in PK parameter estimates, study duration, and study size. The latter two would constitute study budget.

In the subsequent sections we describe the rationale for the investigation we performed to determine the efficiency of Phase 1 designs that address the above issues.

29.6.1   Rationale

In addition to the fact there is only scanty information on the design of FTIH studies in nononcology indications, there is the need for a knowledge base in order to choose among competing designs for the conduct of a FTIH study in a nononcology indication.

29.6.2   Objective

To develop a process to evaluate the comparative efficiency of FTIH study designs, one must take into account mild adverse events (AEs), dose proportionality, the precision in PK parameter estimates, study duration, and study size. A mild adverse event was chosen because most AEs experienced in FTIH studies are mild (see Section 29.5). A desirable design(s) should be able to appropriately characterize the exposure–AE response curve (for the incidence of a mild AE).

29.6.3 Methodology

29.6.3.1 Assumptions

1. A new molecular entity exhibiting one-compartment pharmacokinetics with first-order absorption was assumed. The typical (mean) values of the population PK parameters for the NME were $1\,h^{-1}$, $17.5\,L/h$, and $50\,L$ for absorption rate constant (K_a), apparent clearance (CL/F), and apparent volume of distribution (V/F), respectively. An intersubject variability of 45% (coefficient of variation) was assumed for each of these parameters, and this was assumed to be lognormally distributed with a mean of zero. A proportional error model was assumed for the residual error of 15%.

2. Drug exposure was assumed to be proportional to dose.

3. Mild adverse event (MAE) was assumed to be exposure related.

4. Starting and ending doses were assumed to be the same for all designs.

5. Dosing occasions were restricted as much as possible to 8, unless the nature of the design did not permit the ending dose, which was assumed to be the same for all designs (see Assumption 4 above), to be reached.

6. Total sample size was restricted to 12, unless the nature of the design did not permit and Assumptions 4 and 5 had to be met.

7. All designs were assumed to be single dose-escalation designs.

29.6.3.2 Study Designs

Base Design   To be able to achieve the objective of the investigation, a study design had to be chosen as the base design against which other study designs were compared. The base design is shown in Figure 29.1. The dose levels are indicted by

Panel #	Subject #	Occasion							
		1	2	3	4	5	6	7	8
1	1	P	1X	2X	4X				
	2	1X	P	2X	4X				
	3	1X	2X	P	4X				
	4	1X	2X	4X	P				
2	5					P	8X	16X	32X
	6					4X	P	16X	32X
	7					4X	4X	P	16X
	8					8X	16X	32X	P
	9					P	8X	16X	32X
	10					4X	P	16X	32X
	11					4X	8X	P	16X
	12					8X	16X	32X	P

FIGURE 29.1   Base design (Design 1) with dose levels indicted as X-fold. If, for example, the starting dose is 25 mg (denoted as 1X), then the dose level denoted as 32X is 800 mg.

Doubling Scheme

Dose level	f_n#	Multiples of f_{n-1}	Increase (%)
1	1		
2	2	2	100
3	4	2	100
4	8	2	100
5	16	2	100
6	32	2	100

Fibonacci Scheme

Dose level	f_n#	Multiples of f_{n-1}	Increase (%)
1	1		
2	2 = 1+1	2	100
3	3 = 1+2	1.5	50
4	5 = 2+3	1.67	67
5	8 = 3+5	1.6	60
6	13 = 5+8	1.63	63
7	21 = 8+13	1.62	62
8	34 = 13+21	1.62	62

Modified Fibonacci

Dose level	f_n#	Increase (%)
1	1	
2	2	100
3	3.3	65
4	5	52
5	7	40
6	9	29
7	12	33
8	16	33
9	21	33
10	28	33

FIGURE 29.2   Different dose-escalation schemes with f_n # being the fold increase from the first dose.

X-fold. For example, if the starting dose is 25 mg (denoted as 1X), then the dose level of 32X is 800 mg.

Having chosen the base designs, it was then possible to address study design attributes that enabled us to generate other study designs.

Study Design Attributes

DOSE-ESCALATION SCHEME   Alternative designs were constructed based on three dose-escalation schemes: dose doubling, modified Fibonacci, and the "mixture", which combined dose doubling and modified Fibonacci to be more aggressive in the lower dose range and to be more conservative in the high dose range (e.g., see Figure 29.2). The modified Fibonacci scheme was borrowed from oncology, where it has been used for more than two decades. The Fibonacci scheme is a number sequence, wherein the next number equals the sum of the two previous numbers (1, 2, 3, 5, 8, 13, 21, 34, . . .) may be familiar, but the modification (modified Fibonacci) used in Phase 1 oncology trials (modified Fibonacci) to give decreasing increases ($2n$, $3.3n$, $5n$, $7n$, $9n$, $12n$, $16n$) as multiples of the initial dose, or 100%, 65%, 52%, 40%, 29%, 33%, 33% increase over the previous dose) was introduced by Schneiderman (24). In proposing the modified Fibonacci approach to avoid rapid dose escalation to toxicity in Phase 1 oncology trials, Schneiderman stated, in part: "A decreasing step suggestion also has been made. This is due to Bellman [25, p. 342] in another context, and I have not seen it in any published account of preliminary dose finding [26]. The Bellman suggestion is a form of Fibonacci search. Three decisions have to be made here: the initial dose d, the maximum possible dose d', and N, the number of steps allowable in moving upward from dose d to dose d'. By taking a Fibonacci series of length $N + 1$, inverting the order, and spacing the doses in proportion to the N intervals in the series, one would take smaller and smaller steps in moving from d to d'. This cautious approach has considerable appeal." We have adopted the modified Fibonacci dose-escalation scheme in some of the designs investigated in this study.

	design #				
	doubling	mixture	M.Fibonacci	occasions	subjects
no panel overlay	1, 2	3			
2-panels overlay	5			8	
3-panels overlay	8	9			
sequential panel			4	12	12
alternate panel			6		
			7	9	
dose crossover within panel	14		13	12	
other	15			9	18
subject dosed only once	10			7	28
		11		9	36
			12	10	40

FIGURE 29.3 Design summary showing the characteristics of the different designs. Doubling = dose doubling; mixture = a combination of dose doubling and modified Fibonacci; M. Fibonacci = modified Fibonacci.

OTHER DESIGN ATTRIBUTES   Additional design attributes included panel staggering, study size (number of subjects), dose count, dose frequency per subject, inclusion of a "leading dose" in the randomization scheme, inclusion of a placebo treatment within each dosing occasion and in some designs a subject served as his/her own placebo control, fixing the starting and ending doses, and limitations of the total number of dosing occasions (duration).

Overview of Designs   Fifteen designs were initially considered. However, only nine designs—Designs 1, 3, 5, 6, 7, 9, 10, 14 and 15—were chosen for subsequent investigation. The nine designs were selected on the basis of them being representative of all the design attributes above. The characteristics of the designs are summarized in Figure 29.3, and Figure 29.4 provides examples of a cross section of the designs investigated. In implementing the designs a week's washout period would be allowed before a dose escalation. This is to avoid carryover effects.

29.6.4   Mild Adverse Effect (MAE) Model

A logistic regression model was used to characterize a drug-induced MAE to be observed in the simulated studies using the above designs. Because the exact shape of the exposure–MAE was unknown, assumptions were made about the steepness of the slope of the logistic exposure–MAE curve. Thus, three slope (β) values—0.75, 1.0, and 1.25—were assumed to cover a spectrum of possible slope values ranging from gradual, through moderate, to steep (Figure 29.5). The metric of exposure used was the area under the plasma concentration curve (AUC). The area under the exposure–MAE curve (AUC-AE) was used as the metric for quantifying the exposure–MAE relationship. The code for implementing the logistic regression model is in Appendix 29.1.

(A) Study Design 3 (Fixed Sequence Nonoverlapping Panel Design-Mixture)

Panel #	Subject #	Occasion							
		1	2	3	4	5	6	7	8
1	1	P	2X	4X	8X				
	2	1X	P	4X	8X				
	3	1X	2X	P	8X				
	4	1X	2X	4X	4X				
	5	1X	4X	P	8X				
	6	P	2X	8X	11X				
2	7					11X	P	20X	27X
	8					11X	15X	P	32
	9					11X	15X	20X	P
	10					15X	P	27X	32X
	11					P	15X	27X	32X
	12					11X	20X	27X	P

(B) Study Designs 5 and 9 (Fixed Sequence Overlapping Panel Designs)

Study Design 5 (Doubling with Two-Panel Overlap)

Panel #	Subject #	Occasion							
		1	2	3	4	5	6	7	8
1	1	P	1X	2X	4X				
	2	1X	P	2X	8X				
	3	1X	2X	P	4X				
	4	1X	2X	4X	P				
2	5			P	4X	8X	16X		
	6			4X	P	8X	16X		
	7			2X	8X	P	16X		
	8			2X	4X	8X	P		
3	9					P	8X	16X	32X
	10					4X	P	16X	32X
	11					8X	16X	P	32X
	12					8X	16X	32X	P

Study Design 9 (Mixture with Three-Panel Overlap)

Panel #	Subject #	Occasion							
		1	2	3	4	5	6	7	8
1	1	1X	P	4X	8X				
	2	1X	2X	P	8X				
	3	P	2X	4X	8X				
2	4		4X	P	11X	15X			
	5		2X	8X	P	15X			
	6		2X	4X	11X	PX			
3	7				P	11X	15X	20X	
	8				8X	P	15X	20X	
	9				8X	11X	15X	P	
4	10					P	20X	27X	32X
	11					11X	P	20X	32X
	12					11X	20X	27X	P

FIGURE 29.4   Selected study designs representative of the types of designs investigated. Design 3 is a fixed sequence, crossover, nonoverlapping panel design. The dose-escalation scheme in Design 3 is a mixture of dose doubling and modified Fibonacci. Designs 5 and 9 are fixed sequence, crossover, overlapping panel designs. The dose-escalation scheme in Design 5 is dose doubling, while that in Design 9 is a mixture of dose doubling and modified Fibonacci. Design 6 is an alternating crossover design with a modified Fibonacci dose-escalation scheme. Study Design 10 is a sequential parallel dose design with a modified Fibonacci dose-escalation scheme.

(C) Study Design 6 (Alternating Crossover Design with Modified Fibonacci)

Panel #	Subject #	Occasion											
		1	2	3	4	5	6	7	8	9	10	11	12
1	1	P			3.5X			11X			27X		
	2	1X			5.5X			P			27X		
	3	1X			P			11X			32X		
	4	1X			5.5X			15X			P		
2	5		1X			5.5X			P			27X	
	6		P			5.5X			15X			32X	
	7		2X			8X			15X			P	
	8		2X			P			20			32X	
3	9			2X			8X			20X			P
	10			3.5X			8X			P			32X
	11			P			11X			27X			32X
	12			3.5X			P			20X			27X

(D) Study Design 10 (Sequential Dose Design with Modified Fibonacci)

Panel #	Subject #	Occasion									
		1	2	3	4	5	6	7	8	9	10
1	1	1X									
	2	P									
	3	1X									
	4	1X									
2	5		2X								
	6		P								
	7		2X								
	8		3.5X								
3	9			3.5X							
	10			P							
	11			3.5X							
	12			5.5X							
4	13				5.5X						
	14				P						
	15				5.5X						
	16				8X						
5	17					8X					
	18					P					
	19					8X					
	20					11X					
6	21						11X				
	22						11X				
	23						P				
	24						15X				
7	25							15X			
	26							15X			
	27							20X			
	28							P			
8	29								20X		
	30								P		
	31								20X		
	32								27X		
9	33									27X	
	34									27X	
	35									32X	
	36									P	
10	37										32X
	38										32X
	39										32X
	40										P

FIGURE 29.4 *Continued*

FIGURE 29.5   Relationship between drug exposure and mild adverse event (MAE) based on a logistic regression model with slope ranging from 0.25 to 1.5.

29.6.5   Simulation and Data Analysis

With the assumed population PK parameters, the PK model, and the exposure–MAE relationship, a clinical trial simulation was performed in S-Plus (Insightful Corporation, Seattle, WA). Two thousand four hundred profiles were simulated for each design and analyzed in S-Plus. Dose proportionality was estimated using the power model (26) and mixed effects modeling in S-Plus. Population PK parameter estimates were obtained using nonlinear mixed effects modeling in S-Plus.

   Study designs were compared by assessing the quality of PK parameter estimates, the resulting safety profile, the duration of the trial, and the assumed budget required to perform that design. A design or designs that could best describe the adverse exposure–response relationship was preferred, taking into account the other factors.

29.6.5.1   *Design Efficiency–Cost Metrics*
New metrics were proposed to evaluate design efficiency. The design efficiency–cost metric had to account for the wide variety of designs incorporating different dose-escalation schemes and the different dose levels within each dosing occasion. The overall efficiency–cost metric included these and other factors investigated. To arrive at the efficiency–cost metric, it was assumed that each efficiency measure contains mean (M) and standard deviation (S) with I components of efficiency measures in total. A two-stage linear weighted cost function, which addressed normalization and weighting for each component, was developed:

$$\text{Efficiency–cost metric} = \sum_{i=1}^{I} a_i \left(b_{i1} M_i + b_{i2} S_i \right)$$

where

$$\sum_{i}^{I} a_i = 1 \quad \text{and} \quad b_{i1} + b_{i2} = 1$$

The weighting for the ith component of efficiency–cost measure is denoted by a_i, if three components are considered in the efficiency–cost metric, for example, the set of (a_1, a_2, a_3) contains weighting factors for the precision of PK parameter estimate, dose proportionality, and safety estimate, respectively. Within the ith component, b_{i1} is the weight assigned to the mean property of the ith component of efficiency measure, and b_{i2} is the weight assigned to the associated standard deviation. For example, if the main objective is to focus on the central location of the ith efficiency–cost measure, then more weight would be assigned to b_{i1} than b_{i2}. The efficiency–cost metric was used in evaluating the performance of the study designs.

29.7 STUDY OUTCOME

29.7.1 Dose Proportionality and Precision in the Estimation of PK Parameters

The designs performed similarly in characterizing dose proportionality (Figure 29.6). This was not unexpected since dose proportionality was assumed in the simulation. The model parameters were estimated to a similar degree of precision by each of the designs. This further supported the confirmation of dose proportionality (Table 29.1). The linearity of the pharmacokinetics can be observed from the nature of the disposition phase of the concentration–time profiles shown in Figure 29.7. The code for simulating the PK profile is in Appendix 29.2 and 29.3.

29.7.2 MAE Characterization

When the designs were rated in terms of their ability to predict the exposure–MAE relationship, three designs performed the best for three categories of steepness of the exposure–MAE investigated. Figure 29.8 shows examples of the rank order of

FIGURE 29.6   A density plot of slope values from fitting a mixed effects power model for the determination of dose proportionality to area under the plasma concentration–time curve (AUC)–dose data for the selected designs that were investigated. The width of 2 times the interquartile distance was used for smoothing.

TABLE 29.1 Summary of Pharmacokinetic Parameters (mean ± SE)[a] for All Designs

Design	K_a (h^{-1})	CL/F (L/h)	V_d/F (L)
1	1.15 ± 0.20	16.98 ± 2.86	56.26 ± 8.57
3	1.14 ± 0.18	16.96 ± 2.75	56.17 ± 9.19
5	1.14 ± 0.19	16.91 ± 2.72	56.27 ± 9.24
6	1.15 ± 0.19	16.98 ± 2.77	56.42 ± 9.36
7	1.15 ± 0.19	16.96 ± 2.72	56.30 ± 9.06
9	1.14 ± 0.19	16.99 ± 2.69	56.21 ± 9.18
19	1.17 ± 0.19	17.05 ± 2.65	57.05 ± 8.11
14	1.15 ± 0.19	17.07 ± 2.85	56.38 ± 8.91
15	1.14 ± 0.16	16.89 ± 2.59	56.78 ± 7.72

[a]SE, standard error.

FIGURE 29.7 Concentration–time profiles for virtual subjects simulated within Design 1.

the performance of the designs. Although the figure depicts the performance for the cases in which the slope was either gradual or moderate, the pattern was similar for the third category of steepness (not shown). In all cases, Designs 3, 6, and 9 were in the top tier, Designs 1, 5, 7, 14, and 15 in the second tier, and Design 10 (the design with the worst performance) in the third tier.

29.7.3 Effeciency–Cost Metric

Table 29.2 is a summary of the efficiency–cost metric after applying different combinations of different weighting functions to the components of the metric. In

FIGURE 29.8 Box plots of the expected mild adverse event response, E(AUC-MAE), based on two scenarios (i.e., logistic regression slope values of 0.75 and 1) across study designs. The farther it is from the expected line, the poorer the design in charactering the MAE response profile. (A) AUC-MAE–Design relationship (gradual slope, $\beta = 0.75$); (B) AUC-MAE–Design relationship (moderate slope, $\beta = 1$).

generating the weighting functions, greater weight was given to the exposure–MAE response than the precision with which PK parameters were estimated, and dose proportionality. This was necessary because the primary objective of almost every FTIH study is safety and tolerability. Pharmacokinetics and dose proportionality usually constitute the secondary objective.

When adverse effect was the only contributor to the weighting function, the performance of the designs was of the order described in Section 29.7.2 (see Table 29.2, first panel). When no weight was placed on dose proportionality, but weight was placed on precision of PK parameter estimates (weight = 0.1) and safety (weight = 0.9), the outcome was the same (see Table 29.2, second panel). Taking dose

TABLE 29.2 Comparison of Designs Using the Efficiency–Cost Metric with Different Weighting Function Scenarios

Weighting Scenario 1

PK estimate	0
Dose proportionality	0
AE	1

Mean	1
SD	0

Design	Efficiency	Rank
3	0.558	1
9	0.529	2
6	0.512	3
1	0.351	4
5	0.292	5
14	0.291	6
15	0.282	7
7	0.274	8
10	0.000	9

Weighting Scenario 2

PK estimate	0.1
Dose proportionality	0
AE	0.9

Mean	0.8
SD	0.2

Design	Efficiency	Rank
3	0.511	1
9	0.486	2
6	0.467	3
1	0.323	4
14	0.272	5
5	0.266	6
15	0.255	7
7	0.252	8
10	0.000	9

Weighting Scenario 3

PK estimate	0.2
Dose proportionality	0
AE	0.8

Mean	0.8
SD	0.2

Design	Efficiency	Rank
3	0.463	1
9	0.442	2
6	0.421	3
1	0.294	4
14	0.253	5
5	0.240	6
7	0.230	7
15	0.228	8
10	0.000	9

Weighting Scenario 4

PK estimate	0.2
Dose proportionality	0.1
AE	0.7

Mean	0.8
SD	0.2

Design	Efficiency	Rank
6	0.470	1
3	0.412	2
9	0.409	3
7	0.289	4
1	0.268	5
5	0.235	6
15	0.227	7
14	0.224	8
10	0.073	9

proportionality into account in computing the efficiency–cost metric produced a slight alteration in the ordering of the designs in the top tier designs. The rank order was reversed from being Designs 3, 9, 6 to 6, 3, 9.

29.8 DISCUSSION

The similar efficiency of the designs in estimating dose proportionality was not unexpected given the fact that it was an implicit assumption in the simulation study. The power model implicitly assumes that there is a linear relationship between dose and exposure (e.g., AUC).

The designs fell into three tiers based on the safety scenarios considered. Designs (3, 6, and 9) in the top tier described the exposure–AE response curve equally well. Designs 3 and 9 had the mixture dose-escalation scheme with 8 dosing occasions and Design 6 had the modified Fibonacci dose-escalation scheme with 12 dosing occasions. Design 10 severely underestimated the AE response in any safety scenarios considered. This is because for a parallel dose design with less than six active subjects in a cohort, the active events that are detectable with a given power decrease quickly with increasing cohort sizes (18). With Design 10 the cohort size was four. The cohort size had to be kept at four to keep the sample size from being too large. The sample size required to implement Design 10 is 28 subjects. This is a far cry from a sample size of 12 required for the top tier designs (i.e., Designs 3, 6, and 9). The performance of the top tier designs with cohort sizes ranging from four to six can be attributed to the crossover components of these designs. Designs 3 and 9 are fixed sequence crossover designs, while Design 6 is an alternating crossover design.

The performance of the designs as measured by the efficiency–cost metric followed the pattern observed in their ability to predict the safety outcome. The major reason for this is the overwhelming influence of the AE component in the metric. This is rightly so because safety and tolerability are the primary objectives of a FTIH study. In addition, the designs performed similarly in the efficiency with which PK parameters were estimated. The order of the top tier designs was reversed from 3, 9, and 6 to 6, 3, and 9 when dose proportionality was taken into account in the efficiency–cost metric. However, the value of the efficiency–cost metric is similar across these top tier designs (see Table 29.2, panel 4). The slight edge gained by Design 6 is probably due to the efficiency with which dose proportionality is estimated with this design.

Since study duration was another factor of import in choosing designs, Designs 3 and 9 would be preferred over Design 6 because the first two designs have 8 dosing occasions while Design 6 has 12 dosing occasions. Designs 3 and 9 would permit a FTIH study to be completed much quicker for a "go/no go" decision to be made about the progression of the NME in development. By including a 2 week follow-up period, Designs 3 and 9 would take approximately 10 weeks, and Design 6 would take 14 weeks. Thus, Designs 3 and 9 would be preferred over Design 6 because of budgetary considerations.

The second and third tier designs involve larger sample sizes ranging from 18 to 40 (see Figure 29.3) and perform poorly in predicting MAE and their use should not be encouraged.

29.9   SUMMARY

Phase 1 FTIH studies provide an excellent opportunity for learning about an NME's safety, tolerability, and pharmacokinetics (including dose proportionality) in humans. It is important to choose a study design that will enable the primary and secondary objectives of the FTIH study to be met. This would make it possible for the knowledge needed for mission-critical decision making to be extracted from the data collected from such a study.

The issues surrounding the design of FTIH studies are discussed and an investigation on the performance of different FTIH study designs in addressing questions of MAE (an often observed adverse event type in FTIH studies), efficiency of PK parameter estimation, and dose proportionality is described. All designs estimated PK parameters with similar efficiency (precision) and performed similar in characterizing dose proportionality. Given the importance placed on a design being able to predict the exposure–MAE response, the top tier designs (i.e., Designs 3, 9 and 6) were similarly efficient in doing this. An efficiency–cost metric is developed for judging the performance of designs and using the metric. Designs 3, 9, and 6 are ranked the best designs among the designs investigated. Taking budgetary considerations into account, Designs 3 and 9 are preferred over Design 6. Simulation is a useful tool for choosing a design that would address the objectives of a FTIH study.

REFERENCES

1. US Food and Drug Administration, *Critical Path Initiative.* `http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html`. Accessed March 21, 2006.

2. J. L. Lesko, M. Rowland, C. C. Peck, and T. F. Blasche, Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *J Clin Pharmacol* **40**:803–814 (2000).

3. E. I. Ette, V. Garg, and A. Jayaraj, A rational approach to drug development: the exploratory phase. *Clin Res Regul Affairs* **21**:155–177 (2004).

4. W. A. Colburn, Phase I and II trials: designs and strategies. *Drug Inf J* **30**:573–582 (1996).

5. L. M. Haines, I. Perevozskaya, and W. F. Rosenberger, Bayesian optimal designs for Phase I clinical trials. *Biometrics* **59**:591–600 (2003).

6. K. Margolin, T. Synold, J. Longmate, and J. H. Doroshow, Methodologic guidelines for the design of high-dose chemotherapy regimens. *Biol Blood Marrow Transplant* **7**:414–432 (2001).

7. J. Whitehead, Y. Zhou, S. Patterson, D. Webber, and S. Francis, Easy-to implement Bayesian methods for dose-escalation studies in healthy volunteers. *Biostatistics* **2**:47–61 (2001).

8. Y. Zhou, Choice of designs and doses for early phase trials. *Fundam Clin Pharmacol* **18**:373–378 (2004).

9. C. Buoen, O. J. Bjerrum, and M. S. Thomsen, How first-time-in-human studies are performed: a survey of phase I dose-escalation trials in healthy volunteers published between 1995 and 2004. *J Clin Pharmacol* **45**:1123–1136 (2005).

10. J. O'Quigley, M. Pepe, and L. Fisher, Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* **46**:33–48 (1990).

11. M. Winget, Publication and reporting of phase I clinical trials [abstract]. *Control Clin Trials* **16**(3 Suppl 1):S40 (1995).

12. S. F. Dent and E. A. Eisenhauer, Phase I trial design: are new methodologies being put into practice? *Ann Oncol* **7**:561–566 (1996).

13. R. Mick, and M. J. Ratain, Model-guided determination of maximum tolerated dose in phase I clinical trials: evidence for increased precision. *J Natl Cancer Inst* **85**:217–223 (1993).

14. K. T. Perry, From mouse to man: the early clinical testings. *Drug Inf J* **31**:729–736 (1997).

15. S. Senn and E. Farkad, Clinical cross-over trials in phase I. *Stat Methods Med Res* **8**:263–278 (1999).

16. P. C. Boon and K. C. B. Roes, Design and analysis issues for crossover designs in phase I clinical studies. *J Biopharm Stat* **9**:109–128 (1999).

17. Y. Yin and C. Chen, Optimizing first-time-in-human trial design for studying dose proportionality. *Drug Inf J* **35**:1065–1078 (2001).

18. C. Buöen, S. Holm, and M. S. Thomsen, Evaluation of the cohort size in phase I dose escalation trials based on laboratory data. *J Clin Pharmacol* **43**:470–476 (2003).

19. K. E. Sherman, Alanine aminotransferase in clinical practice: a review. *Arch Intern Med* **151**:260–265 (1991).

20. D. E. Amacher, Serum transaminase elevations as indicators of hepatic injury following administration of drugs. *Reg Toxicol Pharmacol* **27**:119–130 (1998).

21. C. Benichou, Criteria of drug-induced liver disorders: report of an international consensus meeting. *J Hepatol* **11**:272–276 (1990).

22. M. Sibille, N. Deigat, A. Janin, S. Kirkesseli, and D. V. Durand, Adverse events in phase-I studies: a report in 1015 healthy volunteers. *Eur J Clin Pharmacol* **54**:13–20 (1998).

23. O. J. Bjerrum, New safe medicines faster: a proposal for a key action within the European Union's 6th Framework Programme. *Pharmacol Toxicol* **86**:23–26 (2000).

24. M. A. Schneiderman, Mouse to man: statistical problems in bringing a drug to clinical trial, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol IV). University of California Press, Berkeley, 1967, pp. 855–866.

25. R. E. Bellman, *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

26. K. Gough, M. Hutchinson, O. Keene, B. Byrom, S. Ellis, L. Lacey, and J. McKellar, Assessment of dose proportionality: Report from Statisticians in the Pharmaceutical Industry/Pharmacokinetics UK Joint Working Party. *Drug Inf J* **29**:1039–1048 (1995).

APPENDIX 29.1   CODE FOR EXPOSURE-AE RELATIONSHIP FIGURE

```
#####################################
## Figure Exposure-AE Response Curve
## b1 is the location parameter (α)
## b2 is the slope parameter (β)
## xx is the exposure

par(mfrow=c(1,1))
xx_seq(0,13,0.1)
b1_-5
b2_c(.25,.5,.75,1,1.25,1.5)
```

```
for(i in 1:length(b2))
#for(i in 1:1)

{

    y_exp(b1+b2[i]*xx)/(1+exp(b1+b2[i]*xx))
    if(i==1)
        {plot(x=xx,y=y,lty=i,ylim=c(0,1),type="l",xlab="Exposure
    Parameter",ylab="Probability of AE Occurance")}
    else
        {lines(x=xx,y=y,lty=i,col=1,lwd=2)}
}

#legend(0, 0.9,as.character(rev(b2)),
    lty=rev(1:length(b2)),col=rev(1:length(b2)),bty="n")
legend(0, 0.98,as.character(rev(b2)),
    lty=rev(1:length(b2)),col=rep(1,length(b2)),bty="n")
text(x=1.2,y=1,"left to right",cex=1.2)
#title(paste("logit model with alpha (",b1,") across different
slopes"),cex=.7)
```

APPENDIX 29.2   FUNCTION CODE FOR PK PROFILE

```
###################################################################
## one compartmental PK model for 1st order absorption and 1st order
##      elimination rates
## ka is the 1st order absorption rate
## k is the 1st order elimination rate
## V is the volume of distribution

## output data of myPK1.s function is ans
## ans is a matrix, which contains rows of level of doses and columns are
## concentrations at all time points
## i.e. dose,t1,t2,t3,t4,.....for each row

myPK1.s_function(ka, k,v,dose,t) # one row per subject/column is
time vector
{
    dose.n_length(dose)
    t.n_length(t)
    ans_matrix(NA,nrow=dose.n,ncol=t.n+1)
    for(i in 1:dose.n)
    {
      ans[i,2:(t.n+1)]_(dose[i]*ka/(v*(ka-k)))*(exp(-k*t)-exp(-ka*t))
    }
    ans[,1]_dose
    ans
}
```

APPENDIX 29.3   CODE OF PK SIMULATION

```
###################################################################
## to simulate n subjects with absorption constant rate 1.7 and
   elm. Rate 0.12
## intersubject variability (cv=50%).
## within subject variability (sd=10%*mean)
## one compartment oral absorption model
# iteration = iter

#** iteration and PK parameter assumptions
iter_10
ka.u_1.7
ka.cv_0.5
ke.u_0.12
ke.cv_0.5
t_c(seq(0.25,1.25,0.25),seq(1.5,3,0.5),seq(4,24,1))
v_30
error.within_0.1 # within subject variability

#** assign the design and the dose multipler
design_design.1 # design data frame
design$dose_design$dose*10 # 1X=10mg

# start
subject_unique(design$subject)
subject_subject[sort.list(subject)]
n.sub_length(subject)

myCol.str_seq(1:length(t))
myCol.str_paste("t",myCol.str,sep="")
myPK.sim_data.frame(matrix(ncol=length(t)+4,nrow=1))
dimnames(myPK.sim)[[2]]_c("subject","dose","iter","occasion",myCol
.str)
n.count_0
for (i in 1:n.sub) # simulation loop for each subject
{
    data1_design[design$subject==subject[i],] # data1 is the
    subset for subject i

    pk.sim.par_data.frame(ka=exp(rnorm(iter,mean=log(ka.
    u),sd=abs(log(ka.u)*ka.cv))),

          k=exp(rnorm(iter,mean=log(ke.u),sd=abs(log(ke.u)*ke.
    cv))))
    for (j in 1:iter)
    {
```

```
    junk_myPK1.s(ka=pk.sim.par$ka[j],k=pk.sim.par$k
[j],dose=data1$dose,v=v,t=t)
    for(k in 1:nrow(junk))
    {
        #to introduce within subject variability
        myError_rnorm(n=length(t),mean=1,sd=error.within)
        junk[k,-1]_exp(log(junk[k,-1])*myError)
        n.count_n.count+1

    myPK.sim[n.count,]_c(subject[i],junk[k,1],j,data1$occasion
    [k],junk[k,-1])

    }

  }

}
```

Design of Phase 1 Studies in Oncology

BRIGITTE TRANCHAND, RENÉ BRUNO, and GILLES FREYER

30.1 INTRODUCTION

Drug development can be seen as "the information-gathering process that ends when the accumulated information is summarized and presented to a regulatory agency for a market-access decision" (1). The usual clinical development, including that of a chemotherapeutic agent, can be thought of as two successive learn–confirm cycles that involve Phase 1 and Phase 2a trials on the one hand and Phase 2b and Phase 3 trials on the other hand (2). Learning and confirming are quite distinct activities, implying different goals, study designs, and analysis modes (see Chapter 8). It is clear that the more efficient the designs of the trials, the more accurate and precise will be the resulting information.

For each phase of drug development, the study design specifies the objectives, the enrollment procedure, the treatment plan over the study, data collection including the definition of primary and secondary endpoints, and data analysis.

The present chapter focuses on the design of Phase 1 studies, as this critical step in clinical drug development has benefited from recent and significant advances in methodology. The design of the trials depends on the drug and on the study. Indeed, the approach used for a cytotoxic drug cannot be used, as it stands, for a cytostatic drug such as targeted therapy. The former is based on decreasing tumor load leading to potential benefit in terms of improvement of survival or at least of quality of life, while the latter may slow or stop tumor growth without a reduction in tumor load. Another difference is the shape of the dose–response curve. For cytotoxic drugs, the shape is usually monotonically increasing (concept of dose–intensity response), while some cytostatic drugs, such as the cytokine interferon-γ, may exhibit a bell-shaped dose–response curve of immunomodularity activity and antitumor efficacy (3, 4). The consequence of these differences is that the endpoints for determining the optimal drug dose will be different for cytotoxic and cytostatic agents for which biological endpoints or pharmacokinetic (PK) endpoints are to be used (5, 6).

The design of studies involving cytostatic agents, which are expected to be less toxic because of their target specificity, is similar to those developed in non-anticancer therapies. For example, at the Phase 1 level, when safety, pharmacology, and toxicology studies performed in animals show low toxicity, healthy volunteers can be recruited. Such Phase 1 studies have been performed for marimastat, a matrix metalloproteinase inhibitor (7), and for erlotinib, a HER1/EGFR tyrosine kinase inhibitor (8). Because such designs are detailed in another chapter of this book, only designs of studies of cytotoxic agents, or of cytostatic drugs when presenting an unacceptable risk such as carcinogenicity, are developed in this chapter.

30.2 OBJECTIVES FOR PHASE 1 STUDIES

Phase 1 clinical trials, and particularly first-time-in-human (FTIH) trials are small, uncontrolled, sequential learning studies designed to:

- Determine the safety and toxicity profiles of a new agent
- Estimate an optimal safe dose for subsequent studies
- Describe the pharmacokinetic (PK) behavior of the drug after its administration

In the case of cytotoxic agents for which the mechanism of action is nonspecific, toxicity is the major endpoint in such studies and dose–response curves can be established. First, the dose-limiting toxicity (DLT) must be defined. It may be a single adverse event (AE) or a combination of toxic events such as "any irreversible grade ≥2 AE" or "grade 4 thrombocytopenia >1 week or neutropenia with WBC < 1000 for >2 weeks or neutropenic fever >1 week."

After defining the DLT, the main objective is to determine the maximum tolerated dose (MTD), which is defined as the dose inducing a given frequency of probability of severe toxicity, and the dose recommended for Phase 2 studies.

30.3 SUBJECT ENROLLMENT FOR PHASE 1 STUDIES

In oncology, participants in cancer Phase 1 trials of cytotoxic agents, or of cytostatic drugs that appear to have a high risk of adverse event(s) such as carcinogenicity, are patients with any type of advanced solid tumor or hematological disease, and for whom all currently available therapies have failed, or are patients for whom there is no effective therapy against the patient's tumor type (9). These patients consent to participate in the trial as a last resort hoping for some efficacy of the agent under investigation. Thus, from an ethical point of view, one should design cancer Phase 1 trials to minimize the number of patients treated at nontherapeutic doses because there is some expectation of effectiveness of the agent.

The inclusion criteria usually require that patients have normal organ function and most often require they are between 18 years old to an upper limit of age. For a long time, this condition excluded children and the elderly from clinical trials. Today, because of the specificity of these populations, Phase 1 trials are performed with some constraints and using the population PK approach.

30.4 TREATMENT PLAN FOR PHASE 1 STUDIES

The treatment plan is defined in order to reach the objectives defined above. The structuring of the design requires one to choose:

- Starting dose
- Way to escalate the doses including the determination of the dose levels
- Number of patients per level
- Patient assignment to a dose level
- Stopping rules

For ethical reasons, the first administered dose in humans must not induce any serious toxicity in any patients. The starting dose level is most often one-tenth the lethal dose in mice if it is nontoxic in dog; otherwise one-third of the toxic low dose in dog is selected. This is not expected to cause any significant toxicity.

The main differences between the diverse designs encountered in drug development are related to the dose-escalation scheme, the number of patients per level, and the stopping rule definition. The oldest and most frequently used dose-escalation method for the last 20 years is the well known "standard" method based on Fibonacci series. Because of the limitations of this method, more sophisticated approaches have been developed, namely:

- Accelerated titration design
- Nonparametric up-and-down designs
- Bayesian designs
- PK-guided dose escalation using allometry information

30.4.1 Standard Dose-Escalation Design

A "standard" Phase 1 trial design involves cohorts of three patients. Thereafter, the dose-escalation scheme is derived from Fibonacci series (10). The principle of these series is that each number is equal to the sum of the two previous numbers. In order to obtain decreasing increases for the levels, clinicians use a scheme based on the so-called modified Fibonacci series with the second level at twice the starting level, and the subsequent levels being 67%, 40%, and 33% greater than the preceding. Usually, the escalation is performed between patients and no escalation is permitted within the same patient. Because these series are slightly different between studies, when reporting the details of the design it would be more accurate to detail the series used rather than saying "according to the modified Fibonacci scheme." Escalation continues until the frequency of a given degree of dose-limiting toxicity meets some constraint criteria. The constraint depends on the type of encountered toxicity—irreversible, life-threatening, or reversible and manageable toxicity.

Generally, the escalation design is as shown in Figure 30.1.

FIGURE 30.1   Flowchart of the standard dose escalation.

- If no dose-limiting toxicity occurs in the three patients of the cohort, escalation continues.
- If a dose-limiting event occurs in one of the three patients, three additional patients are included at the same dose level. If none of the three additional patients presents toxicity, escalation continues; otherwise the trial stops.
- If two or three patients present with dose-limiting event, escalation stops.

Once stopping the escalation, three more patients are included at the dose preceding the stopping dose.

The MTD is defined as the level preceding the stopping dose and is often the recommended dose for further Phase 2 studies.

There is no compelling scientific basis for the approach, but experience has shown it to be safe. Although this traditional design has been widely implemented during Phase 1 studies in oncology, it presents different issues. The first one is that too many patients receive subtherapeutic doses when the starting dose is far from the MTD. The second problem is that the escalation scheme results in too lengthy trials. A third problem of the standard approach is the lack of precision of the MTD around the actual MTD. In fact, the MTD provided by such trials represents the dose for which the percentage experiencing the DLT ranges from 15% to 70% (11) and may be different from the actual MTD.

A more rational and efficient use of available resources should help to optimize the design of Phase 1 trials leading to a better evaluation of the dose–toxicity relationship and an improved precision of Phase 2 dose recommendation, while minimizing the number of patients included. The improvement should be a more appropriate starting dose, fewer patients per level, and a more appropriate dose-escalation scheme, keeping in mind the safety of the patients.

30.4.2 Accelerated Titration Design

The accelerated titration design (ATD) is a two-stage design that has been developed to overcome some limitations of the standard design (12). The principle is to accelerate, during a first stage, the dose-escalation scheme. Briefly, only one patient is included per dose level until the first instance of DLT occurs, and the dose steps are doubled. When the first instance of DLT or any grade 2 toxicity is observed, the cohort is expanded to 6 or 3 patients, respectively. Then, in a second stage the standard scheme is used with the conventional stopping rules. The intrapatient escalation is permitted as follows: escalate in case of grade 0–1 toxicity at a subject previous course, deescalate in case of grade 3 or greater toxicity.

The authors demonstrated that ATD substantially reduced the number of undertreated patients, because of the reduced number of patients per cohort at the first dose levels, and because intrapatient dose escalation gives a greater chance for patients to receive drug doses more likely to provide antitumor response. Consequently, patients who would be withdrawn from the study, because of tumor progression, may be able to remain longer in the study. This approach results in a substantial increase in the information obtained, particularly regarding cumulative toxicity. Another advantage is that ATD shortens the length of the trial by using a more aggressive escalation scheme until the occurrence of a toxic event.

Because ATD is more aggressive than the standard approach, it may be associated with a higher level of risk for the patients. This means that a careful definition of the level of toxicity considered as acceptable must be stated.

A successful implementation of ATD in a Phase 1 study of an analog of paclitaxel has been reported by Plummer et al. (13). The use of ATD led to the generation of a great deal of valuable toxicity information at doses recommended for the subsequent Phase 2 study. Moreover, the majority of the included patients received doses within the therapeutic range, which was linked to a high level of observed antitumor activity.

30.4.3 Nonparametric Designs

Nonparametric approaches have been developed mainly because of the lack of knowledge about the dose–response relationship at the beginning of the Phase 1 trial and because of the small sample size in these trials. Most of them use the up-and-down scheme. Several authors proposed a design based on the "random walk rules" (RWR) (14, 15), which provides an accurate estimate of MTD as a quantile, or the use of isotonic regressions.

30.4.3.1 Random Walk Rules (RWR) Design

Random walk rules design is a generalization of up-and-down designs such as those developed by Storer (16), which target the quantile corresponding to a probability of tolerable toxicity equal to 0.33. Durham et al. (14, 17) proposed a biased coin up-and-down design (BCD) to allocate doses in Phase 1 clinical trials. This method has been modified by Stylianou and co-workers (18, 19) who proposed the accelerated biased coin up-and-down design (ABCD), which deals with the long evaluation periods of the patient. Both are used to find the dose at which the probability

of toxicity is Γ. The goal of BCD is to center the frequency distribution of design doses around a specific unknown quantile.

The scheme is identical to the ATD until the first toxicity is observed. At that point, the RWR design starts with the following escalation rules:

- If a toxic event is observed in patient j at the dose level d_i, then assign patient $j + 1$ to level d_{i-1}.
- If no toxic event is observed in patient j, flip a biased coin with a probability of heads in the range [0, 0.5]. If it lands on the head side, then assign patient $j + 1$ to level d_{i+1}, else to level d_i.

Once one determines the escalating scheme for a rapid attainment of the MTD, the way to estimate the MTD must be chosen. Different estimators of MTD have been proposed, including the empirical mean of the frequency distribution, maximum likelihood estimation (MLE) under a logistic model, and weighted least squares. Because of the small sample sizes that are employed in Phase 1, none of the estimators present good characteristics. Stylianou and Flournoy (19) included an isotonic regression estimator involving linear interpolation and compared the efficiency of this new estimator to the previous ones. The approach assumes that toxicity is nondecreasing with dose and fits an isotonic regression to accumulated data. All the estimators performed similarly, regarding the convergence, since they reach their equilibrium point in about 20 subjects. The precision was better with the MLE and the isotonic estimator. Moreover, the latter ran faster than the other estimators and rarely overestimated the MTD when it was included in the lattice of the fixed doses. This is nicely illustrated by an example in bone marrow transplantation, where Durham et al. (14) showed that using BCD would have provided more useful information, with the same number of patients, than with other up-and-down designs (19, 20), even when starting far from the target.

30.4.3.2 Methods Involving Isotonic Regressions

Another family of nonparametric approaches involves the isotonic regression. Leung and Wang (21) proposed isotonic regression as a means of designing Phase 1 clinical trials. At any point, the assigned dose to a given patient is the dose for which the estimated toxicity is closest to the MTD. The authors presented different stopping rules and recommended a stopping rule that is simple to implement by clinicians: for example, stop the trial if the same dose has been assigned in three consecutive cohorts.

All the above designs consider toxicity as a binary variable. However, in oncology, toxicity is not reported as toxic or nontoxic but according to the World Health Organization (WHO) scales, which are ordinal scales ranging from 0 to 4. Paul et al. (15) have investigated the possibility to include toxicity as an ordinal variable rather than a dichotomous variable. The principle is to define quantiles for each grade of toxicity by establishing the dose–response relationship at the different grades. An acceptability of toxicity grade must be defined taking into account the nature of the toxicity. The quantiles are then estimated using the multidimensional isotonic regression.

30.4.4 Bayesian Designs

In the standard scheme, MTD is identified from the data. Therefore, because no estimation is involved, MTD is a statistic rather than a parameter, which is strongly dependent on the design and sample size used. In Bayesian approaches, MTD is viewed as a parameter of a dose–response curve, which is most often monotonic. MTD is then defined as the quantile of doses corresponding to a prespecified probability of toxicity. Different methods have been proposed such as the continual reassessment method (CRM), which has evolved through several modifications, the escalation with overdose control (EWOC), and the decision theoretic approach (DTA), which may be combined with a Bayesian optimal design (BOD). All these methods have been developed with the objectives of reducing the number of patients receiving ineffective or uninformative dose levels and of obtaining an accurate estimation of MTD. They follow the same scheme and are derived from the CRM. Briefly, the dose assignments for the first patients are made on the basis of prior information obtained from preclinical trials (in vitro and/or animal trials) or any other available trial. Once the patients have been evaluated regarding the toxicity, the posterior distribution is obtained by combining the gained information with the prior information using Bayes's theorem. The subsequent assignments are made on the basis of this posterior distribution and the trial continues until reaching a predefined stopping criterion.

Bayesian designs start with an ordered space of preselected dose levels based on preclinical studies, and believed to cover an acceptable range of DLT probabilities, including the target probability:

$$\mathcal{D} = \{d_1 = d_{\min} < d_2 < \cdots < d_k = d_{\max}\} = \{d_i, \, 0 < d_i < d_{i+1}\}, \quad i = 1, 2, \ldots, k$$

30.4.4.1 Continual Reassessment Method

The continual reassessment method (CRM) introduced by O'Quigley et al. (22) is the best known Bayesian approach and has been widely presented in the literature, either in its original form or with several modifications and extensions such as the bivariate approach.

Original CRM

1. Before the trial begins, a dose–toxicity curve is established from the information gained from previous studies (preclinical, clinical).
   - Toxicity is a single binary variable, toxic or no toxic response.
   - The dose–toxicity model, $\Psi(d_i, \theta)$, $i = 1, 2, \ldots, k$, is built assuming that the probability of observin g a toxic event θ at each dose level d_i is a monotonic function increasing with dose. O'Quigley et al. (22) chose the one-parameter power model, $p_i = \Psi(d_i, \theta) = [F(d_i)]^\theta$, where F is any distribution function. Other examples of such a function may be the hyperbolic tangent model, $p_i = \Psi(d_i, \theta) = [(1 + \tanh d_i)/2]^\theta$, or the one-parameter logistic model used by many authors (19, 23–26):

$$p_i = \Psi(d_i, \theta) = \frac{\exp(a_0 + \theta d_i)}{1 + \exp(a_0 + \theta d_i)}, \quad i = 1, 2, \ldots, k \tag{30.1}$$

where the intercept a_0 is fixed and θ is a model parameter to be estimated, assuming a prior distribution. The unit exponential distribution ($g(\theta) = \exp(-\theta)$) has been used extensively, or a gamma distribution. Simulations performed by Chevret (24) showed that the form of the prior distribution did not significantly affect the results.

2. The second step is the determination of the target probability of DLT, usually 20% of risk of unacceptable toxicity or in a range of 20–33%. MTD is defined as the $100 \times \pi$ percentile of the dose–toxicity relationship.

3. The first patient is assigned to receive a starting dose for which the probability of toxicity is closest to the target, 0.20, for example. Once toxicity outcome is evaluated from this patient, the dose–toxicity curve is updated by computing the posterior distribution of θ using Bayes's theorem as follows:

   - If we denote $\Omega_j = \{(d_1, y_1), (d_2, y_2), \ldots, (d_j, y_j)\}$, where d_m ($d_m \in \mathcal{D} = \{d_1 = d_{min} < d_2 < \cdots < d_k = d_{max}\}$) and y_m are, respectively, the dose and the observed response of patient m ($m = 1, 2, \ldots, j$), we can summarize all the information available after the evaluation of patient j by the posterior distribution, $f_{\Omega_j}(\theta)$, with

$$\int_0^\infty f_{\Omega_j}(\theta)\, d\theta = 1, \quad j = 1, 2, \ldots, n \tag{30.2}$$

   - For the first patient, $f_{\Omega_1}(\theta)$ is $g(\theta)$, the prior distribution of the parameter of the dose–response curve.
   - For the $(j + 1)$th patient, the posterior distribution is updated using the following Bayesian calculation:

$$f_{\Omega_{j+1}}(\theta) = \frac{f_{\Omega_j}(\theta)\,\Phi(d_j, y_j, \theta)}{\int_0^\infty f_{\Omega_j}(x)\,\Phi(d_j, y_j, x)\,dx} = \frac{g(\theta)\prod_{m=1}^{j}\Phi(d_m, y_m, \theta)}{\int_0^\infty g(x)\prod_{m=1}^{j}\Phi(d_m, y_m, x)\,dx} \tag{30.3}$$

   where $\Phi(d_j, y_j, \theta) = [\Psi(d_j, \theta)]^{y_j}[1 - \Psi(d_j, \theta)]^{(1-y_j)}$.

4. The posterior probability of toxicity is then used for assigning the dose at which the next patient will be treated with a probability of toxicity closest to the target DLT. The process continues until the maximum number of patients has been reached, usually 20–25 patients. The dose that results in this level of toxicity is defined as the recommended dose for further Phase 2 studies.

Modified CRM   Because of ethical considerations, several modifications have been proposed with the objectives of reducing the risk for a patient to receive a toxic dose and of shortening the length of the trial. They concern the starting dose, the dose-escalation scheme, the stopping rules, and the parameterization of the dose–response model using more than one parameter.

Regarding the starting dose, it appears that the original CRM tends to allocate higher doses compared with the standard design (27). The suggestion is to give the lowest available dose (d_{min}) as the starting dose; thereafter the escalation is per-

formed without dose skipping until the first toxicity is observed, with one patient per level. Some authors, such as Goodman et al. (25), proposed the use of cohorts of more than one patient per level. From this point, cohorts of three patients are included at each level at the same time.

Different stopping rules have been proposed, one of them being that the CRM algorithm ends after a given number of patients have been assigned to the same dose (23). O'Quigley and Reiner (28) proposed to stop the process when the dose that produced the target probability of a DLT is determined with a 90% confidence interval, or until the maximum number of patients has been reached. This modified design has been implemented for a Phase 1 study of intraperitoneal topotecan, a topoisomerase inhibitor (29). The method of dose escalation was efficient and allowed a rapid identification of a dose level of $3\,mg/m^2$, which presented a 20% probability of producing DLT, grade 4 granulocytopenia in this case. Other examples of successful use of this method are reported by Eisenhauer et al. (30).

Because there are situations where the actual MTD is not included in the available dose range, Thall and Russell (31) suggested stopping the trial if it appears that no dose satisfies both safety and efficacy requirements. This stopping rule is based on the posterior probability of too high toxicity for each dose level. Other criteria have been explored by Zohar and co-workers (32, 33) in order to detect early a nonadequate choice of dose range or a prefixed gain in the estimation of probability of response associated with the MTD. To this aim, the authors developed a two-stage design for Phase 1 studies. The first stage is similar to the design implemented by Thall and Russell using stopping criteria based on posterior probability of DLT and on CRM. In a second stage, they continue the trial by including patients at the estimated MTD and stop the trial when there is not any significant gain (measured via a gain function) to be obtained from trial continuation. This second stage focuses on reliability in estimates using Bayesian predictive gain functions.

Because of the reluctance of some investigators to use the Bayesian approach, O'Quigley and Shen (34) have developed a more traditional sequential design according to the likelihood-based estimation, which does not require Bayesian prior distributions. The principle is that the equation for obtaining $f_{\Omega_{j+1}}(\theta)$ is close to the likelihood. The final recommended dose levels using this setting differed little from those obtained with the original CRM (35).

Bivariate CRM (bCRM)   Most of the approaches described previously determine the maximum tolerable dose while ignoring efficacy. Several authors proposed an extension of the CRM to a bivariate design in which the MTD is based jointly on both toxicity and efficacy. A way for doing this extension proposed by Thall et al. (36) is to combine the binary indicator of toxicity (0 = no toxicity, 1 = toxicity) with the ordinal indicator of response (0 = no response, 1 = moderate, 2 = important), into a single multinomial variable Y. The issue here is that combining in a unique variable toxicity and response makes it impossible to estimate the effect of dose upon each outcome separately. Therefore, Braun (37) suggested a model where the probability of toxicity and the probability of response are parameterized separately. As in CRM, a regression model for dose effect is established considering that for each dose d_i in the lattice $\mathcal{D} = \{d_1 = d_{min} < d_2 < \cdots < d_k = d_{max}\}$, there is a probability of toxicity p_{1i} and a corresponding probability of disease progression p_{2i}. The model is then a function that increases with regard to toxicity and decreases with regard to

progression. Once this function has been defined, a bivariate distribution for toxicity and progression is selected, keeping in mind the biological plausibility.

The implementation of this design consists in:

- Including the first n patients at the dose level that is most likely close to the MTD.
- Computing the posterior probabilities for each dose level p_{1i}^n and p_{2i}^n.
- Assigning the next cohort of patients to receive the dose

$$dose = \underset{i=1,2,\ldots,k}{\text{Min}} \sqrt{\sum_{j=1}^{2} w_j \left(p_{ji}^n - p_j^* \right)^2} \tag{30.4}$$

where p_1^* and p_2^* are, respectively, the desired rates of toxicity and progression; and w_1 and w_2 are values between 0 and 1, with $w_1 + w_2 = 1$. This constitutes a generalization of the univariate CRM when w_1 or w_2 is equal to 0.

The bCRM is an attractive approach but presents some problems, one of them being that both toxicity and response must occur in a close temporal proximity. A modification that could deal with situations where response occurs significantly later than toxicities would be of interest in oncology.

Another interesting characteristic of the bCRM is that it enables the CRM to be applied to two groups of patients (38), for example, heavily pretreated versus not pretreated. It is of particular interest when the accrual of patients is insufficient in one of the subgroups, which could take advantage of information gained in the other group.

Time-to-Event CRM (TITE-CRM)   An extension of the CRM has been proposed by Cheung and Chappell (39) and adapted by Braun et al. (40). This method takes into account the duration of the follow-up of each patient as a proportion of the maximum duration of follow-up observed. Data from subjects presenting no toxic event are weighted by that proportion, and data from subjects with toxicity are given full weight. These weights applied to the likehood used in CRM allow the determination of a maximum tolerated cumulative dose (MTCD).

30.4.4.2 *Escalation with Overdose Control (EWOC)*

The dose escalation with overdose control (EWOC) is a Bayesian approach similar to CRM. It is a dose-escalation scheme based on controlling the probability of overdosing a patient and not on targetting toxicity between 20% and 30% of the MTD, as in the original CRM (41, 42). This method, like CRM, sequentially modifies the dose–response curve by including the information of all the patients previously included in the trial, but in this case, the dose–effect relationship deals with a two-parameter model, which can be considered as a tolerance function between two bounds ($d_{\min}$ and $d_{\max}$):

$$F(\beta_0 + \beta_1 d) = \text{Prob}\{\text{DLT}|\text{Dose} = d\}, \quad \text{with } d_{\min} \leq d \leq d_{\max} \tag{30.5}$$

where β_0 and β_1 are unknown parameters constrained to ensure that probability of toxicity is monotonically increasing with dose levels.

At the time of each dose assignment, as for CRM, the posterior cumulative distribution function of MTD is computed by using all the information available. The information includes the doses administered to the previous patients, the covariates, and the highest toxicity levels observed in each patient.

The second difference compared to CRM is the addition of a constraint on overdosing. Using the EWOC approach, the first patient will receive a dose considered as safe, as in standard scheme: $d_1 = d_{min}$. The subsequent doses will be determined from the conditional probability that Γ for a patient k is an overdose given the available information $I_k : \Pi_k(\Gamma) = \text{Prob}\{\text{MTD} \leq \Gamma | I_k\}$. The dose level is selected for each patient with the constraint that the predicted probability that the dose exceeds the MTD is less than or equal to a fixed value of α. That means that the dose level d_k of the kth patient will be selected such as $\Pi_k(d_k) = \alpha$.

Low values of α lead to a cautious dose escalation, when higher values imply a more aggressive escalation. At this stage, the uncertainty of MTD is wide and a low value of α is recommended in order to ensure safety to the patient. As the trial goes on, for the subsequent patients the information increases and the uncertainty of MTD decreases, allowing α to increase in the absence of unacceptable toxicity. The escalation stops when the convergence criterion is obtained.

EWOC requires only the information currently available at the time of inclusion, whereas in the standard scheme the inclusion of a new patient can only be done after evaluation of all the patients from the previous cohort. This results in shortening the trial duration.

EWOC has been used successfully to design a Phase 1 clinical trial in patients with solid tumors receiving 5-fluorouracil (5-FU) in combination with fixed doses of leucovorin and topotecan (41). Cheng et al. (43) used this approach during a Phase 1 clinical trial performed in patients with advanced non-small-cell lung cancer receiving PNU-214936, a superantigen-based immunotherapy. Whereas the activity of this drug is moderated by the neutralizing capacity of anti-SEA antibodies (anti-SEA Ab), the MTD depends on the level of anti-SEA Ab in each patient. The EWOC approach allowed the determination of the MTD, taking into account baseline plasma anti-SEA Ab levels, and the establishment of a dosing algorithm to optimize dose individualization in patients according to their tolerance.

30.4.4.3 *Decision Theoretic Approach or Bayesian Optimal Design*
A disadvantage of the CRM is that as the sample size increases, the doses selected for assignment tend to cluster around the target, most often 20% of risk to develop an unacceptable toxicity (TD_{20}). Consequently, it gives accurate estimates of TD_{20}, but the whole dose–toxicity curve is not described (44). Whitehead and Brunier (45) developed a Bayesian decision approach to estimate the whole dose–toxicity curve while focusing on a given target dose. The approach is a combination of the CRM and EWOC method. EWOC involves a two-parameter logistic regression model, but in log dose, with a prior distribution that is updated with ongoing knowledge from cohorts of three patients. The assignment of doses to give to next patients is fixed by maximizing a gain function while maintaining acceptable risk to the patient, taking into account the safety constraint. This gain function can be compared to the dose selection used in CRM or EWOC, where each patient is assigned to a dose at which the posterior probability of toxicity is close to the target TD_{20} or MTD. In the decision theoretic approach, the gain function is a variance gain. That means the

procedure selects the doses that minimize the asymptotic variance of the maximum likelihood estimator of MTD. A continuous range of possible doses is available, and the gain from investigation is measured in terms of statistical information gathered (45, 46).

Haines et al. (47) suggested including the criterion Bayesian D-optimality, which maximizes some concave function of the information matrix, which in essence is the minimization of the generalized variance of the maximum likelihood estimators of the two parameters of the logistic regression. The authors underline that toxicity is recorded as an ordinal variable and not a simple binary variable, and that the present design needs to be extended to proportional odds models.

30.4.5 Pharmacokinetic/Pharmacodynamic (PK/PD) Guided Dose Escalation Using Allometric Information (PGDE)

It is widely recognized that all the information gained from preclinical studies may help in building more informative and efficient clinical trial designs. However, there is a class of information obtained during animal studies which is most often not applied to the construction of Phase 1 studies, namely, the PK/PD information. This information is of great interest and value in determining the Phase 1 clinical design.

The PK/PD guided dose escalation (PGDE) proposed by Collins et al. (48) is based on the assumptions that:

- Interindividual variability and, more generally, interspecies variability in toxicity are for the most part due to differences in the behavior of the drug after its administration and differences mainly in metabolism, elimination, and binding to endogenous components such as proteins.
- Similar plasma concentrations result in the same biological effect in animal and human.

Collins et al. (48) reported, from retrospective studies in cancer agents, that the area under the concentration–time curve (AUC) at the LD_{10} in mice was similar to the AUC observed at the MTD in human. This AUC value has been suggested to be used as a target to guide dose escalation in Phase 1 studies. Thus, the authors proposed the following scheme:

- Determine AUC at LD_{10} in mice: this is now the target AUC in humans.
- Use the same starting dose as in the standard scheme.
- Determine AUC in the first patients entered in the study.
- Define an escalation scheme in three or four steps scaling between this AUC value and the target AUC.
- The trial stops when either the target AUC or the MTD is reached.

This approach has been implemented in studies of some anticancer drugs. One example is the pharmacokinetically guided administration of melphalan performed in patients with advanced ovarian adenocarcinoma (49). The schedule involved a fixed dose on day 1 (7.9 mg) followed by a second dose on day 2, calculated on

the basis of PK data to achieve a target area under the concentration–time curve (AUC). The study showed that the maximum tolerated AUC, based on grade 3 or 4 hematological toxicities, ranged from 86 to 112 mg·min/L. Another example is that of CI-958, a DNA intercalator (50), where the authors estimated that this trial using PGDE design required 15–18 fewer patients than with the standard Fibonacci design.

However, other examples showed less success (51–53). The reason is that the method cannot be used when there are interspecies differences in drug sensitivity, in metabolism, and in plasma protein binding. An improvement in the use of the preclinical data allowed bypassing some of these limitations, as reported by several authors (54–57).

When developing a new therapeutic compound, relevant toxicokinetic studies are performed in different species such as mice, rabbits, dogs, and monkeys. The results of such studies lead to interspecies scaling in PK parameters based on the assumption that there are physiological and biochemical analogies between mammals, which can be expressed mathematically by allometric equations (58). The allometric approach is based on the power function, where the PK parameters Y are plotted against the body weight W of the different species:

$$Y = aW^b \tag{30.6}$$

The coefficient a and exponent b of the allometric function are parameters to be estimated. This relationship has been refined by including some correcting factors that take into account the specificity of the drug under study. Boxenbaum (59) suggested scaling the data with respect to the maximum life-span potential of each species, in order to remove the biologic clock dependency. Other correcting factors have been used, such as brain weight for drugs metabolized by the mixed function oxidation system (54–56), or renal correcting factors (54–56, 60), or protein binding factors. Mahmood (61) has shown that, depending on the number of correcting factors included in the equation, three or more species are needed for an accurate prediction of PK parameters in human from animal data. These correcting factors should lead to a decrease in the number of failures when using the PGDE approach.

Besides animal studies, in vitro studies give information about the doses and concentrations that show antitumor activity and concentration–response relationships can be established. Knowing the therapeutic range of concentration from the concentration–response relationship, and being able to predict the behavior of the drug in humans from the allometric equations, an optimal design may be implemented for the FTIH study.

Cosson et al. (54) provided another improvement for allometric scaling by using the PK/PD population approach. This approach enables one to use sparse and unbalanced data, which is most often the case in animal studies. Using this approach, they were able to estimate all allometric parameters and all interindividual variabilities in the population and for each species. An example of their code for implementation in NONMEM is presented in their publication (54).

A nice example of using an allometric approach combined with a population data analysis has been prospectively implemented for vinflunine, a third generation of semisynthetic vinca alkaloids, which is an analog of vinorelbine. The authors took

advantage of all preexisting information: preclinical data of vinflunine and data obtained from vinorelbine, which had been widely studied regarding preclinical studies, and PK/PD studies (57, 62, 63). The objective of the analysis was to prospectively predict from animal data the vinflunine (VFL) PK behavior in patients, using a population scaling-up strategy. The goal was to evaluate the putative maximum tolerated dose (MTD) in patients and therefore to propose a rapid dose escalation for the first Phase 1 clinical study aimed at determining the MTD. In a first step they established an allometric equation for vinflunine similar to that previously established for vinorelbine. The NONMEM code is presented in the appendix. The equation obtained took into account body weight (W) and maximum life-span potential (MLP) and allowed the prediction of the PK profile of vinflunine in humans. The PK blood profile and subsequent analysis on the first recruited patient confirmed the profile predicted by the scaling-up approach. The observed terminal half-life and the observed total body clearance were close to the predicted values. Hematological toxicity was the major dose-limiting toxicity (DLT) for vinorelbine in patients and was expected to be also the major DLT for vinflunine. The allometric relationship, the ratio of IC_{50} between vinflunine and vinorelbine obtained from in vitro studies, with a decrease in white blood cell (WBC) count as response, suggested that AUC_{50} for vinflunine should be 5 times greater than the AUC_{50} of vinorelbine. Since the vinorelbine MTD corresponded to an 80% decrease in WBC count, the PK and PD modeling on vinflunine allowed the prediction of the AUC value that would cause a decrease of a target value (80%) in white blood cell number. Then from this AUC value and from the clearance predicted by the model, the dose likely to correspond to the MTD was determined. The PK profiles and responses observed in the subsequent patients were integrated in the model, using the Bayesian approach. This reassessed strategy allowed a secure and rapid determination of the MTD with a dose escalation more rapid than that calculated from the modified Fibonacci suite.

30.5   DATA COLLECTION

Data collection is a crucial step to ensure the success of any trial. Phase 1 is not disease specific. Major endpoints must be precisely defined. They are mainly maximum tolerated dose of the treatment and/or mechanism-based biomarkers in the case of targeted therapy. Data collection is focused on measurement and categorization of toxicities. Usually, the data collected includes, but is not limited to:

- Demographic values, baseline values of the organ functions, as well as values obtained during the follow-up of the patient.
- All the values allowing an accurate evaluation of the endpoints (toxicity and clinical response; biomarkers, surrogate endpoints, or clinical endpoints).
- Treatment data, for example, dose and actual times of start and end of infusion.
- Pharmacokinetic data, with a cautious establishment of the sampling scheme. Most often, since regulatory agencies require model-independent PK parameters during Phase 1 studies, extensive sampling is needed at this stage.

In the case of multicenter studies, the methodology used for the determination of the drug and metabolite levels must be identical between the different centers of investigation, and some crossover validation is highly recommended. The same constraints exist for the estimation of the response (toxicity/efficacy) and for the biological parameters. A typical example is the evaluation of the renal function. For practical purposes, renal function is often estimated using glomerular filtration rate (GFR). This parameter is now essential in monitoring patients on treatment especially for drugs that have a significant amount of elimination by the kidneys. Because GFR is difficult to measure in daily practice, most clinicians estimate GFR using formulas based on a measure of serum creatinine (Scr), age, size, and sex of the patient (64, 65). Other authors developed more accurate equations to estimate GFR from serum creatinine (66, 67). The paper by Wright et al. (67) demonstrated the influence of creatinine assay on the values obtained for serum creatinine and the authors found that the enzymatic assay produced larger serum creatinine values than the kinetic Jaffe assay. Consequently, all the research team would have to decide was which formula and assay technique would be used for determining the biological parameters, such as serum creatinine or creatinine kinase. A review from Tett et al. (68) outlines the different renal elimination pathways and the possible markers that can be used for their measurement.

30.6 PHASE 1 STUDIES FOR A COMBINATION OF TWO DRUGS

In oncology there is the common use of a combination of cytotoxic drugs. Because these combinations may result in interactions between drugs such as synergistic toxicities, Phase 1 studies must be performed to determine the maximum tolerated combination(s) of doses of the administered drugs. In this case, the situation is more complex than with a single agent, because there is not a single MTD, but a set of several possible MTD combinations.

Most often, for each drug, the lattice of dose schedules to explore is established in light of data from previously conducted Phase 1 trials. Several escalation schemes are encountered. For a two-drug combination one could escalate the doses of one drug while using fixed doses for the second one, or escalate both drugs either alternately or at the same time.

Korn and Simon (69) proposed a graphical tool for designing Phase 1 studies in order to target MTD of combinations. They build a tolerable-dose diagram using the information obtained from the single-agent toxicity profiles of each drug. Each type of toxic effect involved (leukopenia, neurotoxicity, etc.) is represented as a weighted combination of the doses d_A and d_B of the two drugs A and B: $d_A + w_j d_B$, where w_j is the relative toxicity of drug B relative to drug A for each kind of toxicity (Figure 30.2). Points above the lines represent combinations leading to unacceptable toxicity. In the case of Figure 30.2, drug A and drug B could be targeted at 400 mg and 20 mg, respectively, without any leukopenia protection. Alternatively, higher doses (e.g., 600 mg and 120 mg for A and B, respectively) could be targeted with adequate leukopenia protection, using G-CSF, for example, and an appropriate escalation scheme. The method can be used with a combination of three drugs with a three-dimensional diagram. Once they established the dose-tolerability diagram,

FIGURE 30.2   Tolerable-dose diagram of drug A and drug B regarding leukopenia, neuro-toxicity, and mucositis.

the authors proposed different escalation schemes, depending on the relative activity of the two drugs and taking into account eventual synergistic toxicities.

Kramar et al. (70) showed that the maximum likelihood CRM previously described in this chapter may be useful in a two-drug combination Phase 1 study. As for a single agent, the method requires a priori dose-toxicity profiles. Isobolograms can be obtained using these profiles available from the monotherapy Phase 1 studies of each drug separately, as well as from previous monotherapy Phase 2/3 studies. The method has been used by Morita et al. (71) for designing a Phase 1 clinical trial of capecitabine in combination with cyclophosphamide and epirubicin.

More recently, Thall et al. (72) proposed an adaptive two-stage Bayesian design for finding acceptable dose combinations for a two-agent Phase 1 study. The application of adaptive designs in two dimensions requires explicit consideration of the toxicity and efficacy surfaces that are a function of dose of each drug and requires that one unmask several critical assumptions, particularly concerning synergy. Wright (73) used a relatively simple model allowing the application of a CRM-style dose-finding scheme in more than one dimension:

$$P(x, y, \theta) = \frac{\exp(a + bx + cy + dxy)}{1 + \exp(a + bx + cy + dxy)} \tag{30.7}$$

where P is a nonlinear function of combined doses x and y, the quadratic form describing interactions between the drug doses. θ is a parameter vector $(\theta_1, \theta_2, \theta_3)$, where the subvectors θ_1 and θ_2 parameterize the two single-agent toxicity probabilities, while θ_3 accounts for interaction between the two agents. This methodology has been applied successfully to a trial combining gemcitabine and cyclophosphamide (72, 73).

30.7   SUMMARY

The standard approach to constructing Phase 1 oncology studies does not fully take advantage of the toxicity experienced by the patients in deciding how to escalate the dose and only focuses on the first one or two cycles without consideration of cumulative toxicity (74). It does not employ the PK/PD knowledge gained from preclinical studies. The starting dose may be viewed as empirical and conservative, in that the dose-escalation process based on the modified Fibonacci series has little scientific rationale, requires too many steps, and exposes too many patients to low and presumably ineffective doses of treatment. The design for Phase 1 trials should be based on methods that increase the probability for the patient to receive therapeutic doses of the agent and also provide a better estimation of, and statistical information on, recommended dose for further Phase 2 trials.

There are several methods that offer the opportunity to shorten trial length, without increasing risk of toxicity for the patients. All the above schemes based on the dose–response curve can deal with exposure variables (peak concentrations, AUC, time above a threshold) or any other meaningful PK parameter in place of dose. Thus, the ideal scheme should be a combination of the existing schemes:

- From in vitro studies determine a target PK exposure.
- From animal studies and studies of analogs of the drug on trial, establish pharmacokinetics–response relationships.
- Escalate on the basis of an exposure–response relationship and not a simple dose–response relationship and using mCRM or likelihood CRM or a nonparametric approach depending on the statistical tools that are available to the team. The response should be based on both toxicity and efficacy separately parameterized according to continuous or at least ordinal variables.

A careful and optimal use of preclinical data is of importance to improve the design of subsequent studies. In this context, the efficiency of clinical trials can be substantially improved by implementation of PK/PD modeling during the trials for several reasons. They allow:

- Identification of agents with saturable clearance mechanisms for which dose escalation will result in unexpected overexposure to the drug.
- Identification of new metabolites not observed in preclinical studies.
- Identification of the impact of different covariates allowing identification of subjects that present a high risk of toxicity.
- Transposition of the results to analogs of the drug, by including the information in new trials.
- Simulation clinical trials in order to optimize the design of subsequent Phase 2 studies.

However, the success of the trial can only be expected if the design is logistically realistic and understandable by the clinicians; and if there is a strong relationship between clinicians, pharmacologists, and statisticians.

ACKNOWLEDGMENTS

REFERENCES

1. L. B. Sheiner and J. L. Steimer, Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu Rev Pharmacol Toxicol* **40**:67–95 (2000).

2. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

3. L. M. Weiner, K. Padavic-Shaller, J. Kitson, P. Watts, R. L. Krigel, and S. Litwin, Phase I evaluation of combination therapy with interleukin 2 and gamma-interferon. *Cancer Res* **51**:3910–3918 (1991).

4. W. C. Kopp, J. W. Smith, C. H. Ewel, W. G. Alvord, C. Main, P. M. Guyre, R. G. Steis, D. L. Longo, and W. J. Urba, Immunomodulatory effects of interferon-gamma in patients with metastatic malignant melanoma. *J Immunother* **13**:181–190 (1993).

5. E. L. Korn, S. G. Arbuck, J. M. Pluda, R. Simon, R. S. Kaplan, and M. C. Christian, Clinical trial designs for cytostatic agents: Are new approaches needed? *J Clin Oncol* **19**:265–272 (2001).

6. K. A. Gelmon, E. A. Eisenhauer, A. L. Harris, M. J. Ratain, and P. Workman, Anticancer agents targeting signaling molecules and cancer cell environment: challenges for drug development? *J Natl Cancer Inst* **91**:1281–1287 (1999).

7. A. Millar, P. D. Brown, J. Moore, W. A. Galloway, A. G. Cornish, T. J. Lenehan, and K. P. Lynch, Results of single and repeat dose studies of the oral matrix metalloproteinase inhibitor marimastat in healthy male volunteers. *Br J Clin Pharmacol* **45**:21–26 (1998).

8. A. Sandler, Clinical experience with the HER1/EGFR tyrosine kinase inhibitor erlotinib. *Oncology (Huntingt)* **17**:17–22 (2003).

9. S. A. Grossman, F. Hochberg, J. Fisher, T. L. Chen, L. Kim, R. Gregory, L. Grochow, and S. Piantadosi, Increased 9-aminocamptothecin dose requirements in patients on anticonvulsants. NABTT CNS Consortium. The new approaches to brain tumor therapy. *Cancer Chemother Pharmacol* **42**:118–126 (1998).

10. G. A. Omura, Modified Fibonacci search. *J Clin Oncol* **21**:3177 (2003)

11. D. Faries, Practical modifications of the continual reassessment method for phase I cancer clinical trials. *J Biopharm Stat* **4**:147–164 (1994).

12. R. Simon, B. Freidlin, L. Rubinstein, S. G. Arbuck, J. Collins, and M. C. Christian, Accelerated titration designs for phase I clinical trials in oncology. *J Natl Cancer Inst* **89**:1138–1147 (1997).

13. R. Plummer, M. Ghielmini, P. Calvert, M. Voi, J. Renard, G. Gallant, E. Gupta, H. Calvert, and C. Sessa, Phase I and pharmacokinetic study of the new taxane analog BMS-184476 given weekly in patients with advanced malignancies. *Clin Cancer Res* **8**:2788–2797 (2002).

14. S. D. Durham, N. Flournoy, and W. F. Rosenberger, A random walk rule for phase I clinical trials. *Biometrics* **53**:745–760 (1997).

15. R. K. Paul, W. F. Rosenberger, and N. Flournoy, Quantile estimation following non-parametric phase I clinical trials with ordinal response. *Stat Med* **23**:2483–2495 (2004).

16. B. E. Storer, Design and analysis of phase I clinical trials. *Biometrics* **45**:925–937 (1989).

17. S. D. Durham and N. Flournoy, Random walks for quantile estimation, in *Statistical Decision Theory and Related Topics*, Vol. V, S. S. Berger and J. O. Berger (Eds.). Springer, New York, 2004, pp. 467–476.

18. M. Stylianou and D. A. Follmann, The accelerated biased coin up-and-down design in phase I trials. *J Biopharm Stat* **14**:249–260 (2004).

19. M. Stylianou and N. Flournoy, Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics* **58**:171–177 (2002).

20. N. Flournoy, A clinical experiment in bone marrow transplantation: Estimating a percentage point of quantal response curve, in *Case Studies in Bayesian Statistics*, C. Gastonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla (Eds.). Springer, New York, 1993, pp. 324–336.

21. D. H. Leung and Y. Wang, Isotonic designs for phase I trials. *Control Clin Trials* **22**: 126–138 (2001).

22. J. O'Quigley, M. Pepe, and L. Fisher, Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* **46**:33–48 (1990).

23. E. L. Korn, D. Midthune, T. T. Chen, L. V. Rubinstein, M. C. Christian, and R. M. Simon, A comparison of two phase I trial designs. *Stat Med* **13**:1799–1806 (1994).

24. S. Chevret, The continual reassessment method in cancer phase I clinical trials: a simulation study. *Stat Med* **12**:1093–1108 (1993).

25. S. N. Goodman, M. L. Zahurak, and S. Piantadosi, Some practical improvements in the continual reassessment method for phase I studies. *Stat Med* **14**:1149–1161 (1995).

26. J. O'Quigley and S. Chevret, Methods for dose finding studies in cancer clinical trials: a review and results of a Monte Carlo study. *Stat Med* **10**:1647–1664 (1991).

27. N. Ishizuka and Y. Ohashi, The continual reassessment method and its applications: a Bayesian methodology for phase I cancer clinical trials. *Stat Med* **20**:2661–2681 (2001).

28. J. O'Quigley and E. Reiner, A stopping rule for the continual reassessment method. *Biometrika* **85**:741–748 (1998).

29. S. C. Plaxe, R. D. Christen, J. O'Quigley, P. S. Braly, J. L. Freddo, E. McClay, D. Heath, and S. B. Howell, Phase I and pharmacokinetic study of intraperitoneal topotecan. *Invest New Drugs* **16**:147–153 (1998).

30. E. A. Eisenhauer, P. J. O'Dwyer, M. Christian, and J. S. Humphrey, Phase I clinical trial design in cancer drug development. *J Clin Oncol* **18**:684–692 (2000).

31. P. F. Thall and K. E. Russell, A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* **54**:251–264 (1998).

32. S. Zohar and S. Chevret, The continual reassessment method: comparison of Bayesian stopping rules for dose-ranging studies. *Stat Med* **20**:2827–2843 (2001).

33. S. Zohar, A. Latouche, M. Taconnet, and S. Chevret, Software to compute and conduct sequential Bayesian phase I or II dose-ranging clinical trials with stopping rules. *Comput Methods Programs Biomed* **72**:117–125 (2003).

34. J. O'Quigley and L. Z. Shen, Continual reassessment method: a likelihood approach. *Biometrics* **52**:673–684 (1996).

35. J. M. Heyd and B. P. Carlin, Adaptive design improvements in the continual reassessment method for phase I studies. *Stat Med* **18**:1307–1321 (1999).

36. P. F. Thall, E. H. Estey, and H. G. Sung, A new statistical method for dose-finding based on efficacy and toxicity in early phase clinical trials. *Invest New Drugs* **17**:155–167 (1999).

37. T. M. Braun, The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials* **23**:240–256 (2002).

38. J. O'Quigley, L. Z. Shen, and A. Gamst, Two-sample continual reassessment method. *J Biopharm Stat* **9**:17–44 (1999).

39. Y. K. Cheung and R. Chappell, Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56**:1177–1182 (2000).

40. T. M. Braun, J. E. Levine, and J. L. Ferrara, Determining a maximum tolerated cumulative dose: dose reassignment within the TITE-CRM. *Control Clin Trials* **24**:669–681 (2003).

41. J. Babb, A. Rogatko, and S. Zacks, Cancer phase I clinical trials: efficient dose escalation with overdose control. *Stat Med* **17**:1103–1120 (1998).

42. S. Zacks, A. Rogatko, and J. Babb, Optimal Bayesian-feasible dose escalation for cancer phase I trials. *Stat Prob Lett* **38**:215–220 (1998).

43. J. D. Cheng, J. S. Babb, C. Langer, S. Aamdal, F. Robert, L. R. Engelhardt, O. Fernberg, J. Schiller, F. Forsberg, R. K. Alpaugh, L. M. Weiner, and A. Rogatko, Individualized patient dosing in phase I clinical trials: the role of escalation with overdose control in PNU-214936. *J Clin Oncol* **22**:602–609 (2004).

44. Y. Zhou, Choice of designs and doses for early phase trials. *Fundam Clin Pharmacol* **18**:373–378 (2004).

45. J. Whitehead and H. Brunier, Bayesian decision procedures for dose determining experiments. *Stat Med* **14**:885–893 (1995).

46. J. Whitehead and D. Williamson, Bayesian decision procedures based on logistic regression models for dose-finding studies. *J Biopharm Stat* **8**:445–467 (1998).

47. L. M. Haines, I. Perevozskaya, and W. F. Rosenberger, Bayesian optimal designs for phase I clinical trials. *Biometrics* **59**:591–600 (2003).

48. J. M. Collins, D. S. Zaharko, R. L. Dedrick, and B. A. Chabner, Potential roles for preclinical pharmacology in phase I clinical trials. *Cancer Treat Rep* **70**:73–80 (1986).

49. D. Y. Ploin, B. Tranchand, J. P. Guastalla, P. Rebattu, F. Chauvin, M. Clavel, and C. Ardiet, Pharmacokinetically guided dosing for intravenous melphalan: a pilot study in patients with advanced ovarian adenocarcinoma. *Eur J Cancer* **28A**:1311–1315 (1992).

50. E. C. Dees, L. R. Whitfield, W. R. Grove, S. Rummel, L. B. Grochow, and R. Donehower, A phase I and pharmacologic evaluation of the DNA intercalator CI-958 in patients with advanced solid tumors. *Clin Cancer Res* **6**:3885–3894 (2000).

51. L. Gianni, Learning from CI-941 about pharmacokinetically guided dose escalation. *Eur J Cancer* **28A**:1302–1304 (1992).

52. M. A. Graham, D. R. Newell, B. J. Foster, L. A. Gumbrell, K. E. Jenns, and H. A. Calvert, Clinical pharmacokinetics of the anthrapyrazole CI-941: factors compromising the implementation of a pharmacokinetically guided dose escalation scheme. *Cancer Res* **52**:603–609 (1992).

53. C. van Kesteren, R. A. Mathot, J. H. Beijnen, and J. H. Schellens, Pharmacokinetic–pharmacodynamic guided trial design in oncology. *Invest New Drugs* **21**:225–241 (2003).

54. V. F. Cosson, E. Fuseau, C. Efthymiopoulos, and A. Bye, Mixed effect modeling of sumatriptan pharmacokinetics during drug development. I: Interspecies allometric scaling. *J Pharmacokinet Biopharm* **25**:149–167 (1997).

55. K. W. Ward, J. W. Proksch, P. D. Gorycki, C. P. Yu, M. Y. Ho, B. D. Bush, M. A. Levy, and B. R. Smith, SB-242235, a selective inhibitor of p38 mitogen-activated protein kinase. II: In vitro and in vivo metabolism studies and pharmacokinetic extrapolation to man. *Xenobiotica* **32**:235–250 (2002).

56. K. W. Ward, J. W. Proksch, K. L. Salyers, L. M. Azzarano, J. A. Morgan, T. J. Roethke, J. E. McSurdy-Freed, M. A. Levy, and B. R. Smith, SB-242235, a selective inhibitor of p38 mitogen-activated protein kinase. I: Preclinical pharmacokinetics. *Xenobiotica* **32**:221–233 (2002).

57. L. Nguyen, P. Variol, and C. Puozzo, Population pharmacokinetic scaling up from animal to patient of a novel anti-cancer drug (vinflunine): a tool to predict the maximum tolerated dose in early clinical trial. PAGE 10, Abst 229, 2001. [`www.page-meeting.org?abstract=229`]

58. H. Boxenbaum, Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J Pharmacokinet Biopharm* **10**:201–227 (1982).

59. H. Boxenbaum, Interspecies pharmacokinetic scaling and the evolutionary-comparative paradigm. *Drug Metab Rev* **15**:1071–1121 (1984).

60. I. Mahmood, Allometric issues in drug development. *J Pharm Sci* **88**:1101–1106 (1999).

61. I. Mahmood, Interspecies scaling of renally secreted drugs. *Life Sci* **63**:2365–2371 (1998).

62. L. Nguyen, B. Tranchand, C. Puozzo, and P. Variol, Population pharmacokinetics model and limited sampling strategy for intravenous vinorelbine derived from phase I clinical trials. *Br J Clin Pharmacol* **53**:459–468 (2002).

63. P. Variol, L. Nguyen, B. Tranchand, and C. Puozzo, A simultaneous oral/intravenous population pharmacokinetic model for vinorelbine. *Eur J Clin Pharmacol* **58**:467–476 (2002).

64. D. W. Cockcroft and M. H. Gault, Prediction of creatinine clearance from serum creatinine. *Nephron* **16**:31–41 (1976).

65. R. W. Jelliffe, Letter: Creatinine clearance: bedside estimate. *Ann Intern Med* **79**:604–605 (1973).

66. A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, and D. Roth, A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of diet in Renal Disease Study Group. *Ann Intern Med* **130**:461–470 (1999).

67. J. G. Wright, A. V. Boddy, M. Highley, J. Fenwick, A. McGill, and A. H. Calvert, Estimation of glomerular filtration rate in cancer patients. *Br J Cancer* **84**:452–459 (2001).

68. S. E. Tett, C. M. Kirkpatrick, A. S. Gross, and A. J. McLachlan, Principles and clinical application of assessing alterations in renal elimination pathways. *Clin Pharmacokinet* **42**:1193–1211 (2003).

69. E. L. Korn and R. Simon, Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. *J Clin Oncol* **11**:794–801 (1993).

70. A. Kramar, A. Lebecq, and E. Candalh, Continual reassessment methods in phase I trials of the combination of two drugs in oncology. *Stat Med* **18**:1849–1864 (1999).

71. S. Morita, M. Toi, T. Kobayashi, Y. Ito, Y. Hozumi, S. Ohno, H. Iwata, and J. Sakamoto, Application of a continual reassessment method to a phase I clinical trial of capecitabine in combination with cyclophosphamide and epirubicin (CEX) for inoperable or recurrent breast cancer. *Jpn J Clin Oncol* **34**:104–106 (2004).

72. P. F. Thall, R. E. Millikan, P. Mueller, and S. J. Lee, Dose-finding with two agents in phase I oncology trials. *Biometrics* **59**:487–496 (2003).

73. J. G. Wright, Adaptive dose-finding in two-dimensions: a phase 1 oncology example. PAGE 13, Abst 542, 2004. [`www.page-meeting.org?abstract=542`]

74. B. C. Goh, Design of phase I and II clinical trials in oncology and ethical issues involved. *Ann Acad Med Singapore* **29**:588–597 (2000).

APPENDIX 30.1 NONMEM CODE FOR ALLOMETRIC SCALING

```
$PROB Final Scale-up model

; Three-compartmental open PK model with first order elimination
; CL= a1*WT^b1.MLP^c1
; Vc= a2WT^b2(1+C2Sex)

$INPUT ID NSUB AMT RATE TIME DV EVID CMT WT SEX RACE BW MLP
$DATA allspec.prn IGNORE=C
$ABBREVIATED DERIV2=NO

$SUB ADVAN11 TRANS4

$PK
TVCL=THETA(1)*WT**THETA(7)*MLP**THETA(8)
; WT is actual body weight
; MLP is the Maximum Life span Potential
CL=TVCL*EXP(ETA(1))

TVV1=THETA(2)*WT**THETA(9)*(1+THETA(10)*SEX)
; Sex=0 if female or 1 if male
V1=TVV1*EXP(ETA(2))

V2=THETA(3)*WT**THETA(11)*EXP(ETA(3))
V3 =THETA(4)*WT**THETA(12)*EXP(ETA(4))
K31=THETA(5)*WT**THETA(13)
K21=THETA(6)*WT**THETA(14)
Q3=K31*V3
Q2=K21*V2
S1=V1

$ERROR
  Y=F*EXP(ERR(1))
  IPRED = F
  IRES = DV - F

$THETA
  (0,5.,10.) ; Intercept of Clearance
  (0,2.5,5.) ; Intercept of Vc
  (0,6.,12.) ; V2
  (0,10.,100.) ; V3
  (0,2.5,5.) ; K31
  (0,0.1,1.) ; K21
  (0,.9,2.) (-.9,-.5,0) ; slopes of clearance with weight and MLP
  (0,.9,2.) (0,.2,.9) ; slopes of Vc for weight and Sex
  (0,.9,2.) (0,.9,2.) (0,.9,2.) (0,.9,2.) ;Exponential terms

$OMEGA
  .25 .25 .25 .25

$SIGMA
  .15
```

Design and Analysis of Clinical Exposure: Response Trials

DAVID HERMANN, RAYMOND MILLER, MATTHEW HUTMACHER, WAYNE EWY, and KENNETH KOWALSKI

31.1   INTRODUCTION AND MOTIVATION

After completing Phase 1, a typical Phase 2a trial would be undertaken to provide confidence the new drug actually imparts the desired pharmacology or drug effect in the target patient population. Subsequently, one or more additional Phase 2 trials would be undertaken to provide sufficient confidence in the risk/benefit to warrant investment in large, confirmatory, Phase 3 clinical trials. While some dose-finding (exposure–response) is possible in Phase 3, it is usually impractical to study more than one or two dose strengths as several hundred patients are often recruited into each treatment arm to definitively demonstrate effect, estimate risk/benefit, and subsequently support regulatory approval. Therefore, it is critical that the Phase 2 program in combination with relevant information collected from preclinical and Phase 1 trials provides adequate information across a wide dose range to ensure the right dosing strategy is selected for Phase 3.

Unfortunately, getting the dosing strategy right in a clinical drug development program appears to be an underappreciated and often overlooked objective of a clinical development strategy. Studying the wrong dose in Phase 3 leads to tremendous inefficiencies in the clinical development process as failed trials will need to be repeated with the right dose. Repeating a Phase 2 or 3 trial most often delays the launch of a new drug by several months or more. Losing six months of market exclusivity translates into hundreds of millions of dollars in lost revenues. Despite the complexity and size of modern clinical development programs, the right dose is all-too-often identified late in development or even after regulatory approval.

In fact, about one in five medicines approved by the Food and Drug Administration (FDA) between 1980 and 1999 had a change in the drug label related to dosage regimen. Most often the labeled dose was decreased due to safety concerns

(1). Interestingly, the likelihood of dose-related label changes did not appear to improve over the 1980s and 1990s, suggesting this problem persists.

The consequences of getting the dose wrong extend beyond inefficient drug development. Importantly, administering the wrong dose places the patient at risk. If the dose is too high the likelihood of adverse events is increased. If the dose is too low the treatment will be ineffective. During the 1980s the benzodiazepine midazolam (Versed) was developed and approved as a shorter acting version of the highly successful diazepam (Valium). Midazolam was approved by the FDA at the end of 1985 and initially marketed at a starting dose of 0.1–0.2 mg/kg for conscious sedation. By September 1988, 86 cases of serious cardiorespiratory events had been reported, 46 resulting in death. Subsequently, the starting dose was decreased severalfold to 0.014 mg/kg (2). Only later was a definitive pharmacokinetic/pharmacodynamic (PK/PD) trial conducted using EEG as a high-resolution measure of drug response. This study, conducted in three volunteers, clearly demonstrated that midazolam was approximately five times more potent than originally thought (3). If the early clinical development plan focused on elucidating the potency and dosage regimen, perhaps the right dosage recommendation could have been put forward in the initial drug label.

In his book, *Overdose, the Case Against the Drug Companies* Dr. Jay Cohen puts forward the opinion that the current clinical drug development process and resulting drug label often fail to provide the physician with the necessary information to optimally treat patients. He summarizes the results of the present day clinical drug development process as follows: "The result is that only belatedly, years or even decades later, do we discover that lower doses are not only effective, but avoid many side effects. Of course, by this time, tremendous damage has been done to people and their families" (4). The problem according to Dr. Cohen is not with the drugs themselves but rather with dosing recommendations, available dose–response information, and available dose strengths. He also highlights the fundamental difference between the objectives of clinical practice and clinical trials. The objective of a clinical trial is to obtain an estimate of the *mean* treatment responses in a cohort of patients and then compare treatment means via statistical hypothesis testing. This focus is fundamentally different from clinical practice, where the physician treats *individual* patients, not the *average* patient. This "one (dose) size fits all" approach can be problematic as the starting dose in the drug label may be higher (or lower) than the dose needed to successfully treat a given individual. While information on effectiveness of lower doses may appear in the clinical trials section of a drug label, it is often excluded from the dosage and administration section. Consequently, the practicing physician is unaware that lower doses may be effective and can be used to successfully treat certain patients. Additionally, it can be problematic for physicians to prescribe doses outside the range deemed effective for the *average* patient due to lack of availability of lower dose strengths.

To summarize, conducting effective dose-finding or exposure–response studies early in clinical development (e.g., preclinical, Phase 1, and Phase 2) is an essential component of an effective clinical development strategy. The creation of an informative drug label is an ethical imperative.

A well designed Phase 1 program can provide biomarker information useful for addressing exposure–response and dosing strategy questions. The design of Phase 1 trials is reviewed in Chapter 32. The purpose of this chapter is to:

- Review the need to address learning questions as opposed to addressing only confirmatory questions
- Review the regulatory guidelines relating to exposure–response
- Describe trial design characteristics and data analysis strategies for various exposure–response trial designs

31.2   EXPOSURE–RESPONSE STRATEGY AND OBJECTIVES

Before considering plausible trial designs, the first critical step of a successful and efficient development program is to clearly identify the specific question(s) to be addressed by the trial. "What do we want to learn in this Phase 2 trial?" When considering a clinical development strategy, it is useful to recognize that clinical trial questions can be classified as learning or confirming (5). A confirmatory question addresses a simple yes/no question such as: "Does the drug work better than placebo?" Such a question is typically formulated as a null hypothesis: "the treatment group mean is no different from the placebo group mean" (e.g., H_0: Treatment = Placebo). Given this objective, a logical and efficient Phase 2a design is to compare one active treatment arm (e.g., maximally tolerable dose) against a placebo arm with as few patients as necessary to reject the null hypothesis, confirming that the drug works better than placebo (or failing to confirm by failing to reject the null hypothesis). Such a trial would typically be labeled as a "Proof of Concept" or "Phase 2a" trial. Designing a trial that exclusively addresses the question, "Does the drug work?" may be logical for truly novel (i.e., untested mechanism of action) or unprecedented (first-in-class) compounds. In this setting, where the mechanism of drug action is untested and the probability of success is low, demonstrating clinical response in a small Phase 2a trial is reasonable prior to embarking on a larger, more extensive Phase 2b exposure–response trial. This hypothetical example begins to illustrate how the trial objective(s) and study design are implicitly linked. Rarely would an actual early clinical trial focus on one simple confirming question; more commonly, trial objectives include both learning and confirming questions.

When considering a Phase 2 strategy for precedented compounds (i.e., mechanism of pharmacologic action proved to be effective), exclusively addressing the simple confirmatory question, "Does the drug work?" in a Phase 2a trial is a waste of time and resources as the confidence in the pharmacologic mechanism is high. In such cases, a Phase 2 trial should more logically emphasize learning questions (but not to the exclusion of confirming questions). Where confirmatory questions are typically of the "yes/no" variety, learning questions are typically of the "how much" variety. Important learning questions for Phase 2b include:

- What is the lowest dose that gives the desired response?
- What is the response at the highest evaluated dose (e.g., maximum effect)?
- What is the lowest dose that gives the minimum acceptable level of effect (e.g., minimum effective dose)?
- What is the highest dose with acceptable tolerability (e.g., maximum tolerated dose)?

- What is the lowest dose that provides a clinically relevant response in at least X% of patients (e.g., what is the starting dose?)
- What is the likelihood that a nonresponder will respond to a dose increase?
- How long should the patient remain on treatment before considering a dose increase?
- How much effect should be expected after X weeks of treatment?
- How quickly does the beneficial or adverse drug effect disappear after stopping drug treatment?
- What is the difference in beneficial and adverse response between once and twice daily regimens?

Historically, clinical trials emphasized confirming objectives over learning objectives (5). Careful consideration of both learning and confirming trial objectives is the first important step when considering an exposure–response trial design.

A common but inappropriate learning objective of an exposure–response trial is to "determine a no effect dose." While this objective sounds quite reasonable on the surface, careful inspection reveals problems. First, for an effective drug, the only true "no effect dose" is zero. Second, the "no effect dose" is not an intrinsic property of the drug; rather it is a function of the design characteristics of the study. To clarify, the "no effect dose" is often defined as the highest dose group that is *not* statistically different from placebo. The ability to differentiate one treatment group from another is a function of the difference in effect between treatment groups, the variability in the effect, and sample size. Lack of statistical differentiation from placebo (e.g., $p > 0.05$, where H_0 = treatment is not different from placebo) should not be interpreted as meaning the treatment has no effect. An appropriate interpretation would be: "under the conditions of the trial we were unable to detect an effect with low-dose treatment." A subtle but very important distinction in interpretation is declaring no effect for a treatment arm versus an inability to detect an effect. By simply increasing the sample size of the low-dose group, a statistically significant effect may be probable. Defining a "no effect dose" is problematic as the answer could differ across clinical studies. For example, 10 mg could be declared the no effect dose based on a small Phase 2a trial, where 5 mg is the no effect dose based on a larger Phase 2b trial. What is the no effect dose—5 mg or 10 mg? Clearly, the concept of defining a no effect dose is flawed in that the answer is dependent on the characteristics of the study design rather than any meaningful characteristic of the drug itself. Alternatively, a more appropriate learning objective is to "estimate the lowest dose that provides a clinically relevant magnitude of drug response." Such a dose is commonly referred to as the MED or minimally effective dose. When composing learning questions it is vital to carefully think through the details, definitions, and analysis strategy involved with addressing the trial objective.

While a confirmatory question can be well addressed by a simple two-arm design (placebo vs. active), learning questions tend to be addressed by more complex designs where many dose groups are studied, repeated measures are collected over time, and PK/PD model-based data analysis methods are applied to characterize the underlying relationship between dose, drug exposure, and response. Unfortunately, the tendency within the pharmaceutical industry is to focus on confirmatory, hypothesis testing trial objectives as opposed to learning trial objectives. The role

of the pharmacometrician is to ensure that the early clinical development strategy addresses important learning objectives by highlighting the importance of answering critical questions about exposure–response.

In summary, the first step in designing an effective clinical development strategy is to contemplate specific questions that must be addressed in the trial. Appreciating the important differences between learning and confirming questions and the relationship to trial design is fundamental to formulating an effective exposure–response trial design. The following sections of this chapter review various regulatory guidances as they pertain to exposure–response, and the strengths and weaknesses of various exposure–response trial designs and data analysis strategies to address both learning and confirming questions.

31.3   REGULATORY GUIDANCES ON EXPOSURE–RESPONSE

The 1962 amendment to the Federal Food, Drug and Cosmetics Act of 1938 introduced the terms of "adequate and well controlled investigations" for drug approval into the wording of the document. The use of the plural tense led to the wide interpretation that at least two well-controlled clinical trials were required for any submission to the FDA. This act was further amended by the Food and Drug Administration Act of 1997 (FDAMA) to allow determination of substantial evidence of effectiveness as required for approval of a new drug to be based on data from one adequate and well-controlled investigation and confirmatory evidence. The use of the term "confirmatory evidence" in this context does not imply that such evidence must come from a trial with confirmatory objectives. Rather, it has been interpreted by many in drug development to indicate that a well-designed exposure–response study with learning objectives could constitute positive evidence for drug approval in addition to a well-controlled, confirmatory, clinical trial. These recent developments underscore the importance of exposure–response trials, not only in drug development and labeling, but also in gaining market access. The *FDA Guidance for Industry: Exposure–Response Relationships—Study Design, Data Analysis, and Regulatory Applications* (6) emphasizes this point with the statement: "A dose response study is one kind of adequate and well controlled trial that can provide primary clinical evidence of effectiveness." A brief description of various regulatory guidances that advocate the use of exposure–response designs and analyses is provided in this section.

31.3.1   *Exposure–Response Relationships—Study Design, Data Analysis, and Regulatory Applications*

This document (6) provides recommendations for sponsors of investigational new drugs (INDs) and applicants submitting new drug applications (NDAs) or biologics license applications (BLAs) on the use of exposure–response information in the development of drugs, including therapeutic biologics. This guidance describes:

1. the uses of exposure–response studies in regulatory decision making,
2. the important considerations in exposure–response study designs to ensure valid information,

3. the strategy for prospective planning and data analysis in the exposure–response modeling process,
4. the integration of assessment of exposure–response relationships into all phases of drug development, and
5. the format and content for reports of exposure–response studies.

The guidance emphasizes that a dose–response study is an adequate and well-controlled trial that can provide primary evidence of effectiveness and that exposure–response information can support the primary evidence of safety and/or efficacy. To achieve these objectives the document supports the use of randomly assigned dose or plasma concentration and pharmacodynamic (PD) response to establish efficacy. The appropriate design depends on the purpose of the study. It is important to make the distinction that dose–response studies can potentially be designed with both learning and confirming objectives. Jonsson and Sheiner (7) discuss the challenges in using exposure–response models to address confirmatory objectives in a well-controlled, confirmatory trial to provide primary evidence of effectiveness and provide recommendations for their use.

Exposure–response studies are typically designed with learning objectives in mind, where mechanistic models are assumed to describe exposure–response relationships, which do not rely on randomization for making treatment (e.g., dose) comparisons. In this setting, efficiency (e.g., reduction in sample size) can be gained by assuming an exposure–response relationship where estimates of the model parameters provide sufficient precision to meet the learning objectives regarding a drug's effectiveness without having to power the study for specific treatment comparisons. The guidance acknowledges that selection of an appropriate model is complex and is usually based on the simplest model possible that has reasonable goodness of fit, and that provides a level of predictability appropriate for its use in decision making. This is consistent with the learning nature of most exposure–response studies where it may be difficult to specify a priori a model that will best describe the exposure–response relationship and only after exploratory model building may we discern the appropriateness of using the model for decision making.

31.3.2 *Dose–Response Information to Support Drug Registration*

This guidance (8) describes the purpose of dose–response information as helping identify:

1. an appropriate starting dose,
2. the best way to adjust dosage to the needs of a particular patient, and
3. the dose beyond which increases would be unlikely to provide added benefit or would produce unacceptable side effects.

The guidance points out that any given dose provides a mixture of desirable and undesirable effects, where no single dose is necessarily optimal for all patients, and that information regarding the population and individual dose–concentration, concentration–response, and/or dose–response relationship is useful for providing

dosage and administration instructions in product labeling. The guidance comments that these instructions should include information about both starting dosages and subsequent titration recommendations, as well as information on how to adjust dose in the presence of intrinsic (e.g., age, gender, race, organ dysfunction, weight, body surface area, ADME differences) and extrinsic (e.g., diet, concomitant medications) factors.

The guidance supports the concept that useful dose–response information is best obtained from trials specifically designed to compare several doses. The guidance further comments on strengths and limitations of various study designs to assess exposure–response. The advantages and disadvantages are discussed for the parallel dose–response, crossover dose–response, forced titration, and optional titration (e.g., titration to response) study designs; nonetheless, it is pointed out that the list is not exhaustive. This is important since it may be incorrectly assumed that the regulatory preference is to encourage the parallel dose–response study since the guidance contains the statement: "A widely used, successful, and acceptable design, but not the only study design for obtaining population average dose–response data, is the randomized parallel, dose–response study with three or more dosage levels, one of which may be zero (placebo)."

The guidance also notes that integration of knowledge of the exposure–response relationship may provide an economical approach to global drug development, by enabling multiple regulatory agencies to make approval decisions from a common database.

31.3.3 *Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products*

This guidance (9) focuses on:

1. when effectiveness may be extrapolated entirely from existing efficacy studies,
2. when one single adequate and well-controlled study supported by information from other adequate and well-controlled studies may be acceptable, and
3. when information from a single multicenter study may be acceptable.

This guidance follows the regulatory evolution starting with the requirement that a drug simply be safe (FDC Act 1938), through the amendment that required a drug to be both safe and effective (FDC Act amendment 1962), to the Food and Drug Administration Modernization Act of 1997 (FDAMA 1997), which directs the FDA to provide guidance on the quality and quantity of evidence required for drug and biologics approval. FDAMA (Section 115) includes the amendment to the act that makes it clear that the agency may consider "data from one adequate and well-controlled clinical investigation and confirmatory evidence" to constitute substantial evidence if the FDA determines that such data and evidence are sufficient to establish effectiveness. This ruling is another step in allowing the combination of strictly objective evidence (i.e., a well-controlled, confirmatory, clinical trial) together with model-based analyses to serve as adequate evidence for drug

approval. Use of model-based analyses has been encouraged for many years as described in these guidances. The ICH E4 guidance (8) stresses the need for high standards of quality for both data and modeling methodology. This high standard of quality is particularly important if these exposure–response studies are intended as supportive evidence to gain market access.

31.3.4 *Ethnic Factors in the Acceptability of Foreign Clinical Data*

This guidance (10) describes how a sponsor developing a medicine for a new region can deal with the possibility that ethnic factors could influence the safety and efficacy of medicines and the risk/benefit assessment in different populations. The guidance establishes a classification system of intrinsic (e.g., genetic polymorphism, age, gender, height, weight, lean body mass, body composition, and organ dysfunction) and extrinsic (e.g., medical practice, diet, use of tobacco, use of alcohol, exposure to pollution and sunshine, practices in clinical trial design and conduct, socioeconomic status, compliance with medication) ethnic factors that can affect safety, efficacy, dosage, and dosage regimen determinations. Ideally, any candidate for global development should be characterized as ethnically sensitive or insensitive during the early phases of clinical development. If the data developed in one region satisfies the requirements for evidence in a new region, but there is a concern about possible intrinsic or extrinsic ethnic differences between the two regions, then it should be possible to extrapolate the data to the new region with a single bridging study. The bridging study could be a PD study or a full clinical trial, possibly a dose–response study. The guidance describes the bridging data package as selected information from the complete clinical data package that is relevant to the population of the new region, including PK data and preliminary PD and dose–response data. The bridging data package may also include supplemental data obtained from a bridging study in the new region that will allow extrapolation of the foreign safety and efficacy data to the population of the new region. The guidance advises that dose–response trials carried out early in the drug development program may facilitate the determination of the need for, and nature of, any required bridging studies.

31.3.5 *Statistical Principles for Clinical Trials*

This guidance (11) is intended to give direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational drug in the context of its overall clinical development. With respect to trials to show dose–response relationships, the guidance emphasizes key objectives, such as the confirmation of efficacy; the investigation of the shape and location of the dose–response curve; the estimation of an appropriate starting dose; the identification of optimal strategies for individual dose adjustments; and the determination of a maximal dose beyond which additional benefit would be unlikely to occur. The guidance suggests that a number of doses should be used including a placebo whenever appropriate. For this purpose, the application of procedures to estimate the relationship between dose and response, including the construction of confidence intervals and the use of graphical methods, is as important as the use of statistical tests. Trial designs are discussed, however, trial designs more appropriate for dose–response studies are

described in ICH E4 (8). This guidance stresses that detail of the planned statistical procedures should be given in the protocol.

31.3.6   *General Considerations for Clinical Trials*

This guidance (12) describes:

1. internationally accepted principles and practices for drug development and the conduct of individual clinical trials,
2. approaches to facilitate acceptance of foreign data, and
3. ways to promote a common understanding of general principles, approaches, and definitions of relevant terms for the drug development and regulatory review processes.

The guidance emphasizes the importance of exposure–response information in all phases of drug development. This information allows optimal design of subsequent studies, permits an understanding of safety and efficacy outcomes, helps establish dosage and dosing regimens, and permits adjustment of the dosing strategy in the presence of intrinsic and extrinsic factors.

31.3.7   *Challenge and Opportunity on the Critical Path to New Medical Products: Innovation or Stagnation*

This FDA document (13) highlights the challenges and opportunities facing researchers, drug developers, and regulators to use more innovative approaches to drug development to combat the increasing costs of drug development. The following excerpt from this document provides the FDA's views on the importance of exposure–response information in drug development.

> The concept of model-based drug development, in which pharmaco-statistical models of drug efficacy and safety are developed from preclinical and available clinical data, offers an important approach to improving drug development knowledge management and development decision making. Model-based drug development involves building mathematical and statistical characterizations of the time course of the disease and drug using available clinical data to design and validate the model. The relationship between drug dose, plasma concentration, bioPhase concentration (pharmacokinetics), and drug effect or side-effects (pharmacodynamics) is characterized, and relevant patient covariates are included in the model. Systematic application of this concept to drug development has the potential to significantly improve it. FDA scientists use, and are collaborating with others in the refinement of, quantitative clinical trial modeling using simulation software to improve trial design and to predict outcomes. It is likely that more powerful approaches can be built by completing, and then building on, specific predictive modules.

In summary, these guidances collectively underscore the importance of well-designed and analyzed exposure–response trials in drug development and regulatory decision making.

31.4   EXPOSURE–RESPONSE TRIAL DESIGNS AND ANALYSIS STRATEGIES

Recently, the FDA issued a guidance document for exposure–response analyses (6). The term "exposure" operatively broadened the document's scope to include patient-specific measures of treatment relative to traditional group measures such as dose regimen. Patient differences in area-under-the-concentration-curve (AUC) or minimum steady-state concentration (for example) might account for a portion of the variability observed in the clinical responses. In this sense, these exposure measures could increase the information content of the analysis.

The conduct of an exposure–response trial can be broken into four principal components—the hypotheses, the design, the analysis, and the conclusion. These components are interrelated—each of the first three is critical in producing an interpretable and valid trial conclusion. If too many hypotheses are considered when planning an exposure–response trial, then (conditional on fixed resources) the available resources might not be optimally allocated for evaluating many of the hypotheses. Thus, the exposure–response trial could yield ambiguous or inconclusive results. Additionally, the design should be tailored to the hypotheses. For example, if a drug exhibits large peak-to-trough fluctuations in concentration, then safety endpoints (such as QT prolongation) should be sampled around times that exhibit the maximum concentrations. Failure to sample appropriately could result in overestimating the drug's safety margin. Finally, analyses that depend on incorrect assumptions can exhibit inflated Type I errors and lead to, for example, a false conclusion of efficacy. Subsequent trials and decision making depend on valid conclusions of the current exposure–response trial. Inadequate trial results can lead to performing subsequent, confirmatory trials, which increase the expense of developing the drug and can delay the registration. Thus, consideration of (a) simple, direct hypotheses, (b) designs tailored to these hypotheses, and (c) appropriate, robust analyses are imperative for planning an informative exposure–response trial.

The following section describes the features of various designs employed in exposure–response trials including parallel group, crossover, pharmacokinetics/pharmacodynamics driven exposure–response, titration, flexible (adaptive), and random concentration-controlled trials.

31.4.1   Parallel Group Trial Designs

Parallel group trials (Figure 31.1) are the most popular design and are implemented in all stages of drug development (Phase 1 to Phase 3). Many parallel group trial

FIGURE 31.1   Depiction of a four-arm, parallel group trial.

results are published in the literature. Parallel group trials tend to be confirmatory in nature, where the focus is on a comparison of efficacy and safety endpoints by treatment group. In these types of trials, patients are randomized to distinct treatment arms. The treatment arms can consist of single or multiple doses, or fixed-time or event-driven dose titration. When patients are studied, the disease state can be acute or chronic, or even progressing as treatment differences can be compared over time.

The relative simplicity of parallel group trials (compared to crossovers) comes at the price of increased sample size. The treatment comparisons are performed between patients. Thus, overcoming large between-patient variability observed in heterogeneous patient populations requires large sample sizes to achieve suitable statistical power. Using ANCOVA methodologies by adjusting for baseline covariates can help increase the power of treatment comparisons. Yet, since each patient only receives one treatment, individual exposure–response assessment and determination of treatment-by-patient interactions is not possible using parallel group designs.

It should be noted that parallel group trials can also be conducted with learning objectives regarding the exposure–response relationship. In this setting, efficiency can be gained by assuming a model where precision of the parameter estimates provides sufficient evidence to demonstrate an exposure–response relationship. Thus, a reduction in sample size can be made since the objective of such trials may focus on learning about the exposure–response relationship rather than on specific treatment (dose) comparisons. While estimation of the population average exposure–response relationship is possible with a parallel group design, simulation studies suggest that a crossover or titration design can provide a more precise and unbiased estimate of the underlying population exposure–response relationship as well as estimates of individual exposure–response (14). Using simulation to quantify the ability of various trial designs to address key trial objectives can be a valuable tool when considering the advantages and disadvantages of trial designs.

The following highlights how a mixed effects regression analysis impacted the understanding of the exposure–response relationship and regulatory decision making. During the development of gabapentin, a treatment for neuropathic pain, complementary analysis strategies were used to support a regulatory approval decision. Two clinical trials were submitted to support an indication for the treatment of neuropathic pain. One trial included placebo and high-dose treatment groups. The other study included placebo, a low dose and a middle dose. Each study included an initial titration phase, where doses were escalated weekly over 3–4 weeks. The confirmatory analysis for these studies was an ANCOVA on the average daily pain score over the last week of treatment. Active treatment groups were compared to placebo. In addition, data from both trials were combined and a complementary nonlinear mixed effects exposure–response regression analysis was undertaken to address key learning questions regarding dose-selection and influence of patient covariates. This analysis utilized all the daily pain score measures as well as data collected during the titration phase of the trial (e.g., a much richer data set relative to that used for the ANCOVA). Patient-specific exposures were imputed based on a population PK model developed from available PK data. Availability of the exposure–response regression analysis allowed regulatory scientists to address questions regarding the magnitude of gabapentin effect at

tested as well as untested doses. The influence of subpopulations on drug response was also readily assessed. The integrated information summarized by the exposure–response model contributed directly to the regulatory approval decision and was reflected in the clinical studies section of the drug label: "pharmacokinetic–pharmacodynamic modeling provided confirmatory evidence of efficacy across all doses." The gabapentin exposure–response analysis of pivotal efficacy data provided complementary insights to a traditional ANCOVA analysis and contributed to a regulatory approval decision.

31.4.2 Crossover Trial Designs

Crossover trial designs (Figure 31.2) are less often used in clinical practice and are implemented typically only in Phase 1 or 2 of development. Examples of these trials, such as bioequivalence and drug–drug interaction, commonly appear in the literature. In fact, numerous journal articles and textbooks are devoted to the formulation and analysis of these designs. In these trials, patients are allocated randomly to study arms (or sequences). Each patient within the sequence receives a treatment followed by one or more treatments; these treatments can be replicated. The treatments can consist of single or multiple doses.

Between each treatment administration period, a "washout" phase is often instituted to allow the effect of the treatment to dissipate. Dissipation of the drug effect is critical. Otherwise, the effect will "carry over" into the next treatment period. An example of a drug-mediated effect that will not carry over after the drug clears the body is inhibition of prothrombin time in the presence of a reversible fibrinogen antagonist. In addition, disease progression should be chronic and stable where the drug does not alter the state of the disease. Drugs that connote even a partial cure can induce a carry-over effect. Carry-over effects can lead to biased estimates of the treatment differences and invalid inference. Judicious choice of the design can allow estimation of some carry-over effects and reduce biases. For example, the complete set of orthogonal Latin squares design (Figure 31.3) is balanced for carry-over effects. More efficient designs than this, such as the Williams designs (Figure 31.4), are available. This increased efficiency is at the cost of increased assumptions about the degree of the carry-over effects. Ultimately, the scientist must decide between the degree of design efficiency and its robustness to potentially invalid assumptions (15).

FIGURE 31.2   Depiction of a two-period, two-treatment crossover design.

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

FIGURE 31.3 Depiction of four-treatment, four-period, Latin square crossover design, where A, B, C, and D represent different treatments.

Sequence	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
1	P	H	L	MH	ML	M
2	L	P	ML	H	M	MH
3	ML	L	M	P	MH	H
4	M	ML	MH	L	H	P
5	MH	M	H	ML	P	L
6	H	MH	P	M	L	ML

FIGURE 31.4 Depiction of six-treatment, six-period, William's Latin square crossover design, where P is placebo, H is high dose, M is medium dose, L is low dose, MH is medium-high dose, and ML is medium-low dose.

The impetus behind implementing crossover designs is that they are typically more efficient than parallel group trials. That is, crossover designs require smaller sample sizes to achieve suitable power for treatment comparison. This efficiency results from within-patient comparisons of the treatments; that is, each patient serves as his/her own matched control. Mixed effects models or ANCOVA analyses are standard methods of analysis. Parameters, which account for sequence and treatment period, are added to the model, since each individual receives more than one treatment. Additionally, some carry-over effects can be estimated, depending on the design. It should be noted that mechanistic models, which assume a specific parametric form for carry-over (e.g., the progression of the response or continued drug effect), could be fit to data from crossover studies as well. Such analyses generally fall under the generalized nonlinear mixed effects model domain.

Crossover designs, however, have their own deficiencies. Assessment of safety endpoints (when they are not known to be directly related to the magnitude of drug concentrations) is difficult; conventions such as intent-to-treat analyses are not available. Moreover, if each treatment period is lengthy, then the time required to complete a crossover study could be prohibitive. In addition, lengthy crossover trials could lead to increased patient dropout prior to the subsequent treatment period, thereby reducing the efficiency of the trial.

As previously stated, carry-over effects can bias the estimates of the treatment differences. Much work has been published, which provides guidance on how to proceed when carry-over effects are possible. Senn (16) discusses and critiques several of these approaches. Ultimately, when choosing between a crossover and parallel group design, the researcher should consider the merits of an efficient and potentially biased analysis versus an inefficient yet relatively unbiased one.

Modeling and simulation can make use of prior information when considering trial design. A notable example assessed parallel group and crossover designs for studying the effects of a muscarinic acid agonist on the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADASCog) (17). The development team

required information on the efficacy and appropriate dose selection for subsequent studies. Previous information from trials using tacrine was combined with investigational Phase 1 information. Initial power calculations suggested that 60 patients (10 per sequence) enrolled in a crossover would provide sufficient power to detect a 3-point difference in the ADASCog scale. The crossover design achieved a substantial reduction in trial costs, since the conventional parallel group trial would have required 400–500 patients. The simulation study also demonstrated that although the carry-over effects downward-biased the treatment effects, the crossover trial was the most robust design for the sample size considered.

31.4.3   Pharmacokinetics/Pharmacodynamics Driven Exposure–Response Designs

Phase 2 is commonly the stage of clinical drug development where exposure–response data is generated to inform dose-selection decisions. However, in certain situations, a rapidly appearing, readily discernible drug response (e.g., biomarker) can be measured following administration of a single dose of study drug. Availability of such a measure of drug response can be extremely valuable to a clinical drug development program, particularly when information on analogous marketed compounds is available. In such cases it is possible to select therapeutic doses based on first establishing the pharmacologic "fingerprint" (18) of a new compound in a Phase 1 exposure–response trial using a biomarker of pharmacologic effect. The second step requires developing a quantitative relationship between the biomarker response and the clinical outcome. It is often assumed that the concentration–response relationship for the biomarker will be similar to the concentration–response relationship for the clinical response. This need not be the case, so using historic data obtained with marketed compounds in the same drug class can provide valuable insight into the relationship between biomarker and clinical response.

This two-step approach to dose-finding was employed during the development of remifentanil, an ultra-short-acting μ-opioid agonist (19). This class of compound is commonly used as an analgesic and in combination with other agents for surgical anesthesia. Different levels of analgesia and anesthesia are required during a surgical procedure. The desired level of opioid effect differs considerably during anesthetic induction, intubation, incision, and organ manipulation. Target opioid concentrations are also dramatically influenced by the addition of other agents such as propofol and inhaled nitrous oxide (N_2O). An empiric clinical development strategy would require numerous clinical studies designed to determine the right remifentanil dose for all possible surgical settings and combinations, a very expensive and time-consuming strategy. To avoid this empiric and expensive clinical development strategy, a decision was made to "fingerprint" remifentanil relative to the marketed compound alfentanil, by undertaking a high-resolution PK/PD trial. A processed EEG waveform was used as a measure of opioid effect. It was well established that administration of a μ-opioid agonist caused a marked change in the EEG waveform from a high-frequency, low-amplitude wave to a low-frequency, high-amplitude wave (delta activity). In this Phase 1 setting, the EEG was used to obtain a definitive PK/PD fingerprint for remifentanil. Figure 31.5 illustrates the PK and PD time course following administration of a short intravenous infusion of study drug. The mean remifentanil EC_{50} for EEG response in healthy volunteers was estimated at 19.9 ng/mL, approximately 30 times more potent than alfentanil.

FIGURE 31.5 Pharmacokinetics and EEG pharmacodynamics following a short infusion of remifentanil.

FIGURE 31.6 Concordance between observed and predicted therapeutic remifentanil concentrations. The solid diagonal line is the line of unity. The vertical solid lines are the range of predicted mean concentrations based on relative potency scaling.

Estimates of target remifentanil clinical concentrations were obtained by scaling the alfentanil concentration–response relationship for clinical responses (e.g., analgesia, surgery with low-dose propofol) by the relative potency estimate (30-fold) obtained in the EEG trial. The concordance between the observed therapeutic remifentanil concentration estimated from Phase 2/3 clinical trials and predicted therapeutic concentrations is depicted in Figure 31.6. In every case the predicted range

overlapped with the observed therapeutic concentration. Such predictions resulted in an efficient and cost-effective clinical development strategy relative to an empiric clinical development approach.

Using a Phase 1 PK/PD trial to accurately predict Phase 2 doses can only be successful if the PK/PD trial is well designed. Obtaining estimates of the remifentanil therapeutic concentrations based on knowledge of alfentanil therapeutic concentrations required an accurate and precise estimate of the potency of the two compounds. Accordingly, the Phase 1 clinical pharmacology trial was designed to characterize the potency of remifentanil relative to alfentanil using the EEG as a measure of opioid effect (19). Therefore, alfentanil was included in the trial as an active comparator. Recall that the goal of the trial was not to simply determine the remifentanil PK/PD characteristics on EEG response, but rather to compare the potency of the new μ-opioid agonist to the marketed product, alfentanil. By doing so, the EEG could be put into context relative to other clinically relevant measures of drug response, allowing predictions of therapeutic remifentanil drug concentrations. To minimize variability and maximize learning about the relative potency of the two compounds, the trial was designed as a crossover. Each healthy volunteer received a short infusion of remifentanil and alfentanil in a randomized manner on two separate study periods. Importantly, the EEG response changed very rapidly (see Figure 31.5). Only 6–7 minutes after discontinuing study drug infusion, the EEG returned to baseline levels. Consequently, a high-resolution PK and PD sampling strategy was undertaken (e.g., every 0.5 min for the 5 min at the beginning and following study drug administration) to ensure that onset and offset of EEG response was captured. Less intense sampling would have limited estimation of the remifentanil PK/PD fingerprint, thereby limiting the subsequent extrapolation to therapeutic drug concentrations. Another important aspect of this study was dose selection. An infusion rate was selected that produced drug concentrations leading to maximal EEG response. So, with a single infusion of study drug, concentration–response data was obtained across the entire dynamic range of EEG response, obviating the need to study different doses (e.g., infusion rates) of study drug.

Obviously, such data is best analyzed using a PK/PD model-based analysis. In this case a standard two-stage PK/PD analysis was undertaken (19). Under this analysis plan, each individual's data was modeled separately and summary statistics were calculated from the set of parameters. Basic pharmacology informs us that a given concentration of drug should elicit a specific level of drug response. Figure 31.7 depicts hysteresis in the concentration–response relationship following study drug administration. At a given concentration, the drug response is not equal and depends on whether the corresponding time point is on the ascending or descending portion of the concentration–time curve. For example, at a concentration of 10 ng/mL, there is little or no effect 2–3 minutes after initiating the study drug infusion; however 3–4 minutes after the end of the infusion, drug concentrations again reach 10 ng/mL and the EEG effect is nearly maximal. This phenomenon can be explained by assuming drug action takes place in a theoretical effect compartment (e.g., brain) that is linked to the central compartment (blood or plasma) by a first-order rate constant (k_{e0}) that accounts for the distribution of drug between the central and effect compartments. This type of drug model is commonly referred to as an effect compartment model (20,21). Correcting for the temporal delay (by estimating k_{e0}) between changes in drug blood concentrations and changes in effect results in col-

FIGURE 31.7   Concentration–response curve suggesting hysteresis.

FIGURE 31.8   Effect site drug concentration–response curve. The thin line reflects the observed effect associated with the theoretical effect site concentration. The thick line reflects the sigmoidal E_{max} model fit to the observed data.

lapsing the hysteresis loop (22). The resulting effect site concentration–response data was then described with the pharmacologically based inhibitory sigmoid E_{max} model (Figure 31.8):

$$E = E_0 - E_{max}C_e^\gamma \big/ \big(C_e^\gamma + EC_{50}^\gamma\big) + \varepsilon \qquad (31.1)$$

where E is the EEG response, E_0 is the effect in the absence of drug, E_{max} is the maximal drug effect, C_e is the effect site drug concentration, EC_{50} is the effect site

concentration that produces half-maximal response (a measure of drug potency), and γ is a steepness parameter. The potency of remifentanil relative to alfentanil was then estimated by calculating the ratio of the respective EC_{50} values. This potency ratio was critical in estimating the likely therapeutic remifentanil target concentrations (18) resulting in an efficient and effective clinical development strategy.

In summary, establishing the PK/PD fingerprint of a new compound by carefully studying the acute time course of drug response on a biomarker (chemical, physical, or image) can provide valuable guidance for dose selection as well as discerning PK and/or PD differences between patient subgroups.

31.4.4 Flexible Trial Designs

Most clinical study designs are fixed. That is, the characteristics of the design (e.g., sample size, number of dose groups, number of subjects per dose group) are established prior to executing the trial and remain fixed throughout, and the resulting data are analyzed only once after completion of the trial. While historically most trials have utilized fixed designs, there is growing interest in the use of flexible designs. Fixed designs often fail when the assumptions used to design the trial are violated. An overly optimistic estimate of the difference in response between treatment groups or a low estimate of variability can result in a failed (significant *p*-value not reached) clinical trial. One advantage of a flexible design is the ability to reset the sample size to avoid conducting an underpowered trial. In addition, flexible designs can provide a faster time to decision, a vital factor in any clinical development program. The following section highlights various flexible trial designs and provides basic guidance as to when such designs should be considered. Flexible designs include (but are not limited to) group sequential designs, randomized concentration-controlled designs, and Bayesian adaptive dose allocation designs.

31.4.4.1 *Group Sequential Designs*
One type of flexible design is the group sequential design (23). Such a design typically includes one or more preplanned interim analyses. Based on the results of the interim analysis, a preplanned action is taken. Most commonly these actions include:

- early stopping for futility (e.g., low probability of demonstrating beneficial response),
- early stopping for success (e.g., high probability of demonstrating beneficial response), and/or
- resetting of the sample sizes based on a revised estimate of variance and/or effect size.

The advantage of such a design is early feedback allowing for early decision making. The ability to make an early go/no go decision is a very attractive feature of a group sequential design. By stopping early, resources used to study ineffective compounds can be minimized. There is also an important ethical advantage

to implementing early stopping rules in a trial. By all accounts it is unethical to prolong the exposure of a patient to a compound that is in all likelihood ineffective and potentially unsafe. A group sequential design with early stopping rules attempts to stop patient exposure to ineffective treatments at the earliest possible moment. Conversely, interim analysis can provide confidence to continue investing in an effective compound. Also, when there is uncertainty (e.g., first-in-class compound, new clinical endpoint) in the estimate of effect size or variance in the primary measure of efficacy, the group sequential design provides a mechanism for resetting the sample size based on interim results, thus minimizing the likelihood of running an underpowered trial. With a traditional fixed design, if assumptions about effect size or variance are amiss, the trial can end up being underpowered or overpowered. So in summary, the potential advantages of a group sequential design include early stopping for either futility or success as well as resizing the trial based on updated estimates of effect size or variance.

A common concern for a group sequential trial utilizing repeated interim ANOVA analysis is an inflated chance of observing a spurious result leading to an inflated Type I (false positive) error rate. Table 31.1 (24) illustrates the impact on the Type I error rate (α) based on the number of interim analyses each controlled at $\alpha = 0.05$.

To combat the problem of inflated Type I error rates, various "alpha-spending" methods have been developed, where "alpha-spending" refers to the statistical penalty incurred due to repeated interim analyses. By applying an appropriate statistical penalty (e.g., Pocock or O'Brien–Flemming alpha-spending rules, lowers the alpha level to <0.05 for each interim analysis) the overall Type I error rate for the trial is maintained at the traditionally accepted 0.05 level. A number of "alpha-spending" methods have been developed and evaluated over the past few decades. Fortunately, software tools such as East® (Cytel Software Corporation) (25) provide a user-friendly environment to simulate clinical trials using various designs, decision rules, and alpha-spending rules. The graphical and tabular results from such simulations provide insight into the performance of various sequential cohort trial designs and alpha-spending rules. Other disadvantages of the group sequential design are the practical aspects of conducting such a trial. Processes must be established to ensure the rapid retrieval, checking, and data entry of key endpoint data. If three interim analyses are undertaken, the data must be retrieved, cleaned, and entered on four separate occasions. Additional resource is needed to

TABLE 31.1   Overall Type I Error Rate as a Function of the Number of Interim Analyses

Number of Analyses	Type I Error Rate
1	0.05
2	0.08
3	0.11
5	0.14
10	0.19
25	0.25

repeat the statistical analysis, generate tables, and write interim reports for each analysis. In addition, an independent data safety monitoring board is usually convened to discuss interim analysis results and review safety data. So the resources needed to conduct a group sequential design (that does not terminate early) can be greater than those required to run a traditional design. Nonetheless, the pressure to make decisions as early as possible and to eliminate failed trials has increased considerably in clinical development. Consequently, interest in group sequential designs is growing.

Group sequential designs should be considered when there is sufficient opportunity (e.g., time) to learn and take action prior to study completion. For example, if all subjects are recruited very quickly (e.g., 1 or 2 months) there is little opportunity and value in undertaking an interim analysis and making midstudy decisions. Consequently, interim analyses are often planned for large, long-term mortality trials where there is ample opportunity to stop early or resize the trial, minimizing the likelihood of a failed trial. Another reason to consider a flexible design is when the estimates of the likely effect size or variability in response are poorly understood. Obvious cases where such uncertainty exists include the evaluation of an unprecedented compound (e.g., first in class) or new clinical endpoint.

31.4.4.2 *Bayesian Adaptive Dose Allocation Design*

Given the increased pressure to discontinue development of ineffective drug candidates early in development, various sequential cohort designs are now being considered for critical Phase 2 proof of concept and dose-finding designs. An example of a flexible design specifically developed to address proof of concept and dose–response questions is the Bayesian adaptive dose allocation (BADA) design (26). As the name implies this type of design employs a Bayesian algorithm to estimate the underlying dose–response relationship. The design includes early stopping rules for either futility or definitive efficacy similar to the sequential cohort design. A unique feature of the BADA design is adaptive allocation of patients to dose groups. Based on interim estimates of the dose–response relationship, a decision theoretic approach is used to preferentially allocate new patients to dose groups that maximize learning about a particular part of the dose–response curve (e.g., ED_{95}). Dose groups that are ineffective have few patients allocated while doses near, for example, the ED_{95} have more patients allocated. With this approach, resources (e.g., drug supply and patients) are not wasted studying ineffective dose groups. If the probability of demonstrating the desired drug response is very low, the trial can be stopped early for futility. Conversely, if the probability of demonstrating a superior drug response is very high, the trial can be stopped early for efficacy reasons. The BADA design attempts to optimally allocate patients to the dose groups that maximize learning and allow for early decision making. Obviously, these are attractive design characteristics for any clinical development strategy.

A BADA design should be considered when there is adequate time to learn and adapt. Moreover, they are particularly well suited for unprecedented compounds where there is little prior knowledge regarding exposure–response. In contrast, for precedented compounds where there is prior knowledge regarding the class of compounds, a fixed dose design with allocation determined based on clinical trial simulations using prior information regarding the exposure–response relationship may outweigh the advantages of a BADA design where an empirical dose–response

model is employed and requires a burn-in period to collect sufficient information to make an informed decision about subsequent dose allocations. If there is an unwillingness to study ineffective dose groups due to excessive cost (e.g., costly drug supplies or patient assessments) or ethical reasons, then a BADA design should be considered. One disadvantage of this design is the need to develop and validate a relatively complicated data analysis and treatment allocation software application. Another disadvantage of this design is that the underlying statistical methodology (Bayesian, decision theoretic approach) is not familiar and readily accepted by some. Nonetheless, the BADA design is an excellent example of how efficiencies in Phase 2 drug development can be achieved.

31.4.4.3 *Randomized Concentration-Controlled Trial*

Another type of flexible design is the randomized concentration-controlled trial (RCCT) (27). As with a fixed design, a patient is randomly assigned to a prespecified treatment arm at the beginning of the trial. With a fixed design a treatment group is typically defined by dose and drug and does not change during the trial. However, with a RCCT, treatment groups are defined by target drug concentrations. For example, a RCCT may include placebo, low-concentration, and high-concentration treatment groups. Doses are adjusted during the trial in individual patients to achieve the prespecified target drug concentration. Accordingly, the RCCT is flexible with respect to the actual dose administered. Such a design should be considered in cases where the variability in observed drug response is primarily due to PK variability. Controlling for drug concentration rather than dose can lead to increased power in such cases. Consider that patients randomized to a high-dose group can have exposures similar to those randomized to a low-dose group, and vice versa, thereby increasing the variability in response associated with a particular dose and limiting power. The RCCT ensures that patients in the high-concentration treatment group actually have high study drug concentrations, increasing the likelihood of observing a drug response by reducing within-treatment (target concentration group) variability in exposure. In addition, a RCCT should be considered in cases where drug response influences drug concentration and the objective of the trial is to obtain an unbiased estimate of the relationship between drug exposure and response. To illustrate, a compound designed to improve cardiac output will also affect blood flow to clearing organs such as the liver and kidney, leading to correlation between pharmacokinetics and pharmacodynamics. Under these circumstances, randomizing based on dose and then undertaking a concentration–response relationship can lead to biased estimates, as poor responders will be more likely to have higher drug concentrations than good responders.

Importantly, undertaking a RCCT is likely to be more expensive and more complicated to implement relative to a typical parallel group trial where treatment groups are defined by dose and drug. Efficient processes from sample collection to determination of drug concentrations to dose adjustment methods must be carefully considered, validated, and implemented. Mechanisms for ensuring blinding while allowing midstudy dose adjustments must also be carefully considered. In the end, trade-offs between added costs must be weighed against the benefits of improved power. Clinical trial simulation can provide useful quantitative comparisons of the improved efficiency of a RCCT over alternative designs, thus aiding decision making.

Analysis of RCCT is obviously well suited to regression analysis (e.g., nonlinear mixed effects models) as concentration is quantitatively linked to clinical response. While treating concentration group as a class variable in an ANCOVA is a viable analysis approach, this analysis method ignores a basic tenet of pharmacology, namely, the causal link between drug concentration and drug response. In other words, response data observed with the low-concentration group carries with it information about response at the high-concentration and vice versa. Again, the use of clinical trial simulation can be a valuable tool to evaluate the performance of a RCCT relative to alternative trial designs and data analysis strategies.

31.4.4.4   Titration Designs

Another form of flexible trial is the titration design. Under a titration design patients are randomly allocated to treatment groups. Then, during the treatment phase of the trial, the dose of study drug is titrated up or down. A trial where titration is required for all patients in a treatment group would be considered a *forced titration* design. Such a design can be very informative for addressing learning questions related to individual dose–response. Such data is amenable to mixed effects regression analysis, where the analysis model would include an exposure–response component as well as interindividual variability in exposure–response. This type of analysis strategy is ideal for learning about the influence of patient covariates (e.g., age, gender, baseline disease severity) on patients' sensitivity to (e.g., EC_{50}—drug concentration that produces half-maximal response) or efficacy of (e.g., E_{max}—theoretical maximal drug response) the study drug. The model can include time-dependent changes in effect (e.g., temporal lag in effect) or disease progression, assuming a placebo group is incorporated into the design.

Another titration design is the *flexible titration* design. Under this design patients randomized to active treatment are titrated to a different dose if warranted. If the patient does not derive the desired level of effect, the dose of study drug can be increased. Alternatively, if the patient experiences adverse events, the dose can be decreased. The problem with such a design is that the observed mean responses across the dose range are inherently biased. Only unresponsive patients would receive high doses of study drug. Consequently, an ANCOVA analysis could lead to biased findings and inappropriate conclusions regarding dose–response. The strength of such a design is that it reflects the actual clinical setting, where dose is adjusted based on the individual's response, not the average response. Unbiased results can be obtained from a flexible titration design by using a mixed effects regression approach. Sambol and Sheiner (28) undertook a nonlinear mixed effects dose–response analysis of data obtained from a flexible titration study of atenolol and betaxolol, where the dose of study drug was increased if the patient did not experience adequate blood pressure control. The model predicted that mean response at high doses differed considerably from the observed mean response. The observed mean response was biased as only nonresponsive patients received higher doses. The model-based analysis provided an estimated effect at high doses based on data from all patients, not just those receiving high doses. A flexible titration design and appropriate regression analysis are well suited to address learning questions such as estimating the probability of unresponsive patients responding to a dose increase.

31.5 SUMMARY

The primary focus of exposure–response trials is to identify the optimal dosing strategy, which in the past has been underappreciated and overlooked as part of the clinical development strategy. Conducting effective dose-finding or exposure–response studies early in clinical development should be an essential component of the clinical development strategy. Appreciating the important differences between learning and confirmatory objectives and their relationship to trial design is fundamental to formulating an effective exposure–response trial design. The regulatory guidances discussed in this chapter underscore the importance of exposure–response trials. Moreover, regulatory agencies are increasingly asking for more information regarding exposure–response relationships and are advocates for using novel designs to obtain such information. In this chapter we discuss different designs that can be used to obtain useful information regarding exposure–response relationships. To be more efficient in drug development and to improve the quality of the regulatory data package and drug label, we must embrace a Phase 2 strategy focusing on collecting high-quality exposure–response information. In the past, clinical development organization often viewed Phase 2 trials as a "mini-Phase 3" (e.g., asking only confirmatory questions) or even worse a barrier to Phase 3 development. A key to efficient clinical development is to view Phase 2 as a fundamental opportunity to learn about the dose–response characteristics of the drug candidate.

REFERENCES

1. J. Cross, H. Lee, A. Westelinck, J. Nelson, C. Grudzinskas, and C. Peck, Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999. *Pharmacoepidemiol Drug Safety* **11**:439–446 (2002).

2. J. Brooks et al., FDA's deficient regulation of the new drug versed. *Congressional Report*, pp. 100–1086, 1988.

3. M. Buhrer, P. O. Maitre, C. Crevoisier, and D. R. Stanski, Electroencephalographic effects of benzodiazepines II. Pharmacodynamic modeling of the electroencephalographic effects of midazolam and diazepam. *Clin Pharmacol Ther* **48**:555–567 (1990).

4. J. S. Cohen, *Overdose, the Case Against the Drug Companies*. Penguin Putman Inc., New York, 2001, p. 7.

5. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

6. *FDA Guidance for Industry. Exposure–Response Relationships—Study Design, Data Analysis, and Regulatory Applications*. FDA, Rockville, MD, 2003.

7. E. N. Jonsson and L. B. Sheiner, More efficient clinical trials through use of scientific model-based statistical tests. *Clin Pharmacol Ther* **72**:603–614 (2002).

8. International Conference on Harmonisation. *Guidance for Industry E4. Dose–Response Information to Support Drug Registration*. CDER, FDA, Rockville, MD, 1994.

9. *FDA Guidance for Industry. Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products*. FDA, Rockville, MD, 1998.

10. International Conference on Harmonisation. *Guidance for Industry E5. Ethnic Factors in the Acceptability of Foreign Clinical Data*. CDER, FDA, Rockville, MD, 1998.

11. International Conference on Harmonisation. *Guidance for Industry E9. Statistical Principles for Clinical Trials*. CDER, FDA, Rockville, MD, 1998.

12. International Conference on Harmonisation. *Guidance for Industry E8. General Considerations for Clinical Trials*. CDER, FDA, Rockville, MD, 1997.

13. FDA. *Challenge and Opportunity on the Critical Path to New Medical Products: Innovation or Stagnation*. FDA, Rockville, MD, 2004.

14. L. B. Sheiner, Y. Hashimoto, and S. L. Beal, A simulation study comparing designs for dose ranging. *Stat Med* **10**:303–321 (1991).

15. B. Jones and M. G. Kenward, *Design and Analysis of Cross-over Trials*. Chapman & Hall, London, 1989, pp. 197–224.

16. S. Senn, The AB/BA crossover: past, present and future? *Stat Methods Med Res* **3**:303–324 (1994).

17. D. J. Hermann, P. A. Lockwood, W. Ewy, and N. H. G. Holford, Use of computer assisted trial design (CATD) for a Phase 2 go, no-go decision. *Clin Pharmacol Ther* **71**(2):57 (2002).

18. T. D. Egan, K. T. Muir, D. J. Hermann, D. R. Stanski, and S. L. Shafer, The electroencephalogram and clinical measures of opioid potency: defining the EEG–clinical potency relationship ("fingerprint") with application to remifentanil. *Int J Pharm Med* **15**:1–9 (2001).

19. T. D. Egan, C. F. Minto, D. J. Hermann, J. Barr, K. T. Muir, and S. L. Shafer, Remifentanil versus alfentanil: comparative pharmacokinetics and pharmacodynamics in healthy adult volunteers. *Anesthesiology* **84**(4):821–833 (1996).

20. C. J. Hull, H. B. H. VanBeem, K. McLLeod K, et al., A pharmacodynamic model for pancuronium. *Br J Anaesth* **50**:1113–1123 (1978).

21. L. B. Sheiner, D. R. Stanski, S. Vozeh, R. D. Miller, and J. Ham, Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to *d*-tubocuraraine. *Clin Pharmacol Ther* **25**:358–371 (1979).

22. D. Verotta and L. B. Sheiner, Simultaneous modeling of pharmacokinetics and pharmacodynamics: an improved algorithm. *Comput Appl Biosci* **3**:345–349 (1987).

23. C. Jennison and B. W. Turnbull, Group sequential methods with applications to clinical trials. Chapman & Hall/CRC, Boca Raton, FL, 2000.

24. R. L. Davis and I. K. Hwang, in *Statistics in the Pharmaceutical Industry*, C. R. Buncher and J. Tsay (Eds.). Marcel Dekker, New York, 1994, p. 269.

25. East, Cytel Software Corporation. `http://www.cytel.com/East/default.asp`.

26. M. Krams, K. R. Lees, W. Hacke, A. P. Grieve, J. M. Orgogozo, and G. M. Ford, Acute stroke therapy by inhibition of neutrophils (ASTIN): an adaptive dose–response study of UK279276 in acute ischemic stroke. *Stroke* **34**:2543–2548 (2003).

27. H. Kraiczi, T. Jang, T. Ludden, and C. Peck, Randomized concentration-controlled trals: motivation, use, and limitations. *Clin Pharmacol Ther* **74**(3):203–213 (2003).

28. N. C. Sambol and L. B. Sheiner, Population dose versus response of betaxolol and atenolol: a comparison of potency and variability. *Clin Pharmacol Ther* **49**:24–31 (1991).

PHARMACOMETRIC KNOWLEDGE CREATION

Pharmacokinetic/Pharmacodynamic Knowledge Creation: Toward Characterizing an Unexplored Region of the Response Surface

ENE I. ETTE and HUI-MAY CHU

32.1   INTRODUCTION

Pharmacokinetic/pharmacodynamic (PK/PD) knowledge creation is the process of building on current understanding of data that is already acquired by generating more data (information) that can be translated into knowledge. It entails the use of (valid) models to synthesize data, estimate inestimable uncertainty, or supplement data for further knowledge acquisition (1, 2).

The intent of knowledge creation is the characterization of unexplored response surface to aid our understanding of drug action. The response surface can be described as three-dimensional. On one axis are the input variables (controllable factors) such as dosage regimen and concurrent therapies. Another axis incorporates patient characteristics, which summarizes all the important ways patients can differ that affect the benefit to toxic ratio (3). That is, the response surface describes the relationship between the therapy and the effects, and how this relationship varies with patient characteristics and time to explain tolerance or sensitivity. For rational drug development and the optimization of individual therapy, this response surface must be mapped for the target population. This shift has occurred because of a concern for maximizing the benefit/risk ratio for individual patients in addition to answering the question of efficacy. Regarding knowledge of the response surface, PK/PD knowledge discovery (1, 4) and creation (1, 2), as described later, greatly improve the precision of this process, which in turn can result in rational drug development with optimized dosing strategies. PK/PD knowledge discovery is the nontrivial process identifying valid, novel, potentially useful, and ultimately understandable patterns in data by characterizing data structure by means of a model (4).

With the PK/PD knowledge discovery process, information (data) is turned into knowledge, while the PK/PD knowledge creation process results in more knowledge generation. Knowledge extracted or created from a clinical trial data set can then be used for decision making. Thus, after the completion of the PK/PD knowledge creation process, a better comprehension is gained about the response surface. This knowledge and comprehension makes wisdom for rational drug development possible, because wisdom (the knowledge and ability to make the right choices at the opportune time) is the final step of good mission-critical decision making. That is, the created knowledge can then be appropriately applied in the design and conduct of appropriate mission-related clinical trials, or the progression of a compound in development.

Knowledge creation is an emerging, interdisciplinary research field that lives at the intersection of computer science (database, artificial intelligence, graphics, and visualization), statistics, and an application domain such as clinical pharmacology in general, and pharmacometrics in particular.

In the sections that follow types of PK/PD knowledge creation, general steps in the knowledge creation process, data supplementation and the motivation for it, data supplementation procedure, nonparametric approximate Bayesian data supplementation method, structure-based multiple supplementation with a motivating example, and implications of the use of multiple data supplementation for the characterization of an unexplored region of the response are described. The emphasis in this chapter is on the use of data supplementation to characterize unexplored region of the response surface. A discussion on data synthesis, the qualitative characterization of the response surface, and the estimation of inestimable uncertainty has been elegantly presented by Williams et al. (1) and Ette and Onyiah (5). Therefore, these approaches are discussed in brief in the sections that follow.

32.2   TYPES OF PK/PD KNOWLEDGE CREATION

32.2.1   Data Synthesis

When there is a considerable amount of information about the drug, synthesizing data into a coherent package that indicates the drug developer has understanding of the pharmacology and, eventually, good control over the therapeutics of the drug provides a means for knowledge creation about the drug being developed. Data synthesis is performed when available knowledge about the drug is used to simulate a clinical trial to explore study outcome when various controllable and uncontrollable factors are varied. This is a knowledge creation process because the objective is to obtain knowledge about the unknown (i.e., unexplored region of the response surface) using valid models. A case in point is the use of clinical trial simulation to investigate the exposure–response relationship in a first-time-in-human (FTIH) study. This involves not only extrapolation of PK/PD from animal to human, but also the exploration of the response surface, hitherto unknown, for a new compound about to be introduced into humans (1). Data synthesis via clinical trial simulation offers the means of generating complex data sets, which may include the influence of prognostic factors, sample size, and dropouts, for testing new competing analysis methods (6, 7). Clinical trial simulation is covered in detail by other chapters (see Chapters 33–35) of this book.

32.2.2 Qualitative Prediction

Physiologically based pharmacokinetic (PBPK) modeling is a modeling approach that lends itself to knowledge creation (1). The result is a model that predicts the qualitative behavior of the experimental time course without being based on it. Refinement of the model to incorporate additional insights gained from comparison with experimental data yields a model that can be used for quantitative extrapolation beyond the range of experimental conditions. That is, the model allows predictions to be made of the kinetic behavior of a drug at various dose levels and routes of administration.

32.2.3 Estimating Inestimable Uncertainty

Parameter estimation without an appropriate assessment of reliability of the estimates yields no confidence in such estimates. Estimation of uncertainty enables the use of such parameter estimates in data synthesis. Embarking on data synthesis (e.g., clinical trial simulation) using model parameter estimates without associated uncertainty or poorly defined uncertainty will produce unreliable outcomes. Sometimes it is impossible to obtain standard errors for population model parameter estimates when small sample sizes are used for population PK/PD modeling. The bootstrap with winsorization has been proposed for the estimation of inestimable uncertainty—standard errors—for population PK/PD parameters that are usually not obtainable using software such as NONMEM because of small sample size (4).

32.2.4 Data Supplementation

Data supplementation deals with the use of models on available data to generate supplemental data that will be used to characterize a targeted unexplored segment of the response surface. The assumptions about models to be used in data supplementation are not as stringent as those required for data synthesis. That is, the use of predictive models is not an absolute necessity (2).

32.3 GENERAL STEPS IN THE PK/PD KNOWLEDGE CREATION PROCESS

PK/PD knowledge creation from a clinical trial data set is a process that can be formalized into a number of steps. In this section we provide a general framework for the steps needed to be taken in the PK/PD knowledge creation process. These steps could vary depending on the type of knowledge creation approach involved. Subsequently, data supplementation—the PK/PD knowledge creation approach of focus in this chapter—is discussed.

Briefly, the steps in the PK/PD knowledge creation process are as follows:

Step 1. Statement of the objective of the PK/PD knowledge creation process.

Step 2. A data set and/or a valid model summarizing the discovered knowledge from a prior PK/PD knowledge discovery process.

Step 3. Performance of knowledge creation (i.e., data synthesis, estimation of inestimable uncertainty, or data supplementation).

Step 4. Analysis of the data generated in step 3.

Step 5. Application of the knowledge created.

Step 6. Communication of the created knowledge.

The objective of the PK/PD knowledge creation process must be clearly defined. With a clear objective in mind, the path chosen for the PK/PD knowledge creation process can be delineated. For PK/PD knowledge creation via data synthesis, valid models are needed. Data synthesis performed using clinical trial simulation requires the use of valid input/output models for the PK/PD knowledge creation process. Data supplementation, on the other hand, requires model assumptions that are not as stringent as the assumptions made when analyzing, the data created by a data supplementation methodology.

Once data synthesis or supplementation is performed, the data must be analyzed for the created knowledge to be extracted. This can be performed using statistical or population PK/PD modeling approaches chosen by the pharmacometrician/pharmacokineticist. There will be variations in steps 3 and 4 of the PK/PD knowledge creation process depending on whether data synthesis, estimation of inestimable uncertainty, or data supplementation will be performed. The application of the created knowledge occurs when the knowledge gained is fed back into the drug development process to aid the understanding of the response surface of a drug under development. Communication is the key to the usage of the product of the knowledge creation process.

The rest of this chapter focuses on the data supplementation approach for PK/PD knowledge creation.

32.4   DATA SUPPLEMENTATION

Data supplementation deals with the supplementation of data to enable the exploration of an aspect of the response surface that may not have been targeted for exploration in a completed trial. It also deals with the supplementation of data in preclinical animal studies where the destructive nature of the sampling design does not permit the construction of individual profiles for inaccessible tissues. A motivating example that deals with a targeted aspect of an unexplored region of the response surface is discussed next to provide clarity on the approach. Data supplementation in the preclinical animal setting is beyond the scope of this chapter.

32.4.1   Motivation for Data Supplementation

The motivation for data supplementation comes from the following:

- After data from a trial has been analyzed, it may become obvious that the dose range explored was limited, and more information (data) would be needed to gain an understanding of the effect of a dose or doses not studied.

- Abrupt cessation of a clinical trial could occur for nonclinical reasons, such as a nonclinical toxicology study finding. In such a situation, not all subjects would have completed the clinical trial—an incompletely observed study. The question arises as to what the responses of the subjects who could not complete the study would have been if the trial was not stopped abruptly. If a solution was found that provided an insight into what the study outcome would have been, the need for repeating such a study once the nonclinical problems are resolved could be obviated.
- Sometimes the clinical trial data do not lend themselves to the traditional PK/PD analyses. Consider a situation in which drugs are administered as combinations in a clinical trial due to their anticipated synergy, but the concentrations of the primary drug driving the effect is unavailable while that of the synergistic drug is available. In such a situation a PK/PD model cannot be developed to characterize the interaction, but there is the need to characterize the effect that could be produced with a different dose of the interactor drug while the dose of the primary drug remains constant.
- A drug may be found, after a clinical trial, to appear to exhibit an inverted U-shaped response, and it is not clear whether a dose not studied in the trial could have produced an effect on the upswing of the dose–response curve that is more effective than a dose in the downswing of the dose–response curve.

32.4.2 Methodology for Data Supplementation

Multiple supplementation (MS) and its modification thereof—structure-based multiple supplementation (SBMS) approach—is proposed as a method for addressing the issue of data supplementation for the characterization of a targeted region (e.g., effect of a dose or dose range) of an unexplored response surface. The MS approach is an adaptation of the multiple imputation methodology used for augmenting data in missing data situations to enable data analysis on a complete data set. First the procedure for performing data supplementation is described, followed by a review of the multiple imputation (MI) methodology, a description of the MS approach, and a discussion of SBMS in the context of a motivating example.

32.4.2.1 Data Supplementation Procedure
The procedure for data supplementation is as follows:

Step 1. Statement of the objective of data supplementation for PK/PD knowledge creation.

Step 2. Performance of PK knowledge discovery.

Step 3. Covariate data synthesis for virtual subjects in the target dose group(s).

Step 4. PK data synthesis for target dose groups.

Step 5. Discovery of hidden knowledge from real data set to which supplemental data will be added.

Step 6. Implementation of a data supplementation methodology (i.e., MS and its modification, SBMS).

Step 7. Discovery and communication of the created knowledge.

Nonparametric Approximate Bayesian Data Supplementation Method   Rubin (8) described a simple method for MI called the approximate Bayesian bootstrap (ABB). This approach makes it possible to generate proper imputation for Y_{mis} with minimal distributional assumptions: procedures for imputation, whether based on explicit (parametric) or implicit (nonparametric) models, ignorable or nonignorable models, that incorporate appropriate variability among repetitions within a model are called "proper." "Ignorable missingness" occurs when the probability of a missing value is not dependent on the value itself, but may depend on the values of other variables in the data set (9, 10). A variety of proper imputation methods based on both explicit and implicit models, including a fully normal model, the Bayesian bootstrap, and the approximate Bayesian bootstrap (ABB), have been studied by Rubin (11). This approach has been adapted for MS, making it possible to generate "proper" supplementation for Y_{supp} with minimal distributional assumptions. To illustrate the ABB approach for MS, consider a collection of n units with the same value of covariates X, where a subjects were observed and $n_{supp} = n - a$ subjects (virtual) with values to be supplemented. The ABB creates M ignorable repeated supplementations for $m = 1, \ldots, M$ as follows: (a) create a new pool of Y_{obs}^* by sampling a values from $Y_{obs} = (y_1, y_2, \ldots, Y_a)$ with replacement, and (b) select a set of n_{supp} possible values from Y_{obs}^*, again with replacement. By drawing n_{supp} supplemented values from a *possible* sample of Y_{obs}^* values rather than from the Y_{abs} values, the ABB approach generates appropriate between-supplementation variability, at least assuming large sample random samples given covariates X. This is akin to the generation of imputation variability, assuming large sample random samples as demonstrated by Rubin and Schenker (12).

Pooling of Estimates   Following the approach used in the MI paradigm, after M supplementations have been created for a data set, they are then analyzed using a standard PK/PD or statistical package. There are now M completed data sets containing the observed values and the supplemented values instead of one. The PK/PD or statistical analysis must be done M times, once on each complete data set. Across M data sets the results will vary, reflecting the uncertainty due to supplemental observations. The M complete data analyses are combined to create one repeated-supplementation inference.

Let $\hat{\Theta}_m$ and U_m, $m = 1, \ldots, M$, be M complete supplemented data estimates and their associated variances for a parameter Θ, calculated from the M data sets completed by repeated supplementations under one model. For instance, $\Theta = \beta$, $\hat{\Theta}_m$ is the least squares estimate of β, and U_m is the weighted residual mean square error. The repeated supplementation estimate of Θ is the mean of the complete data estimates:

$$\bar{\Theta} = \left(\sum_{m=1}^{M} \hat{\Theta}_m \right) \Big/ M$$

There are two components of the variability associated with this estimate: the average within-supplementation variance,

$$\bar{U} = \left(\sum_{m=1}^{M} U_m \right) \Big/ M$$

and the between-supplementation component,

$$B = \left(\sum_{m=1}^{M} \left(\hat{\Theta}_m - \bar{\Theta} \right)^2 \right) \bigg/ (M - 1)$$

The total variability associated with $\bar{\Theta}$ is given by

$$T = \bar{U} + (1 + M^{-1})B$$

Inference can be made using Θ, T, and a distributional assumption. For example, if Θ is a scalar quantity, the approximate reference distribution for interval estimates and significance tests is a t distribution:

$$(\Theta - \bar{\Theta}) T^{-1/2} \sim t_\nu$$

where the degrees of freedom, ν, are given by (12)

$$\nu = (M - 1)\left(1 + r^{-1}\right)^2 \quad \text{with } r = (1 + M^{-1})B/\bar{U}$$

Thus, a $100(1 - \alpha)\%$ interval estimate for $\bar{\Theta}$ is

$$\bar{\Theta} \pm t_{\nu, 1-\alpha/2} \sqrt{T}$$

The between-subject and within-subject ratio, r, estimates the population quantity $\gamma/(1 - \gamma)$, where γ is the fraction of information about Θ supplemented.

32.5 STRUCTURE-BASED MULTIPLE SUPPLEMENTATION: A MOTIVATING EXAMPLE

This example illustrates how knowledge can be created using a combination of data synthesis, structure revelation, and multiple supplementation (MS) techniques. Since data supplementation was performed based on the structure revealed from the data, as discussed later, this modification of the MS approach is termed structure-based multiple supplementation (SBMS). A parallel dose efficacy study of a drug in development was performed with three dose levels—placebo, 200 mg, and 600 mg. Subjects were sampled for population PK and efficacy analysis. The objective of this PK/PD knowledge creation investigation was to determine a likely treatment outcome if subjects were randomized to a 100 mg dose group that was not studied in an already completed trial. The 100 mg dose group is hereafter referred to as the target dose group. Prior to performing PK/PD knowledge creation, PK knowledge discovery (6) was performed. Thirty-five subjects were administered the test drug on a three times daily basis for 28 days. Eighteen and seventeen subjects were randomized to receive 200 and 600 mg of test drug orally. These subjects provided 974 concentrations, yielding an average of 27.8 concentrations/subject over a 28 day period. The mean (SD) age, weight, and height of subjects were 45 (6.3) years, 85 (31.3) kg, and 171.6 (12.3) cm, respectively. There were 21 male and 13 female subjects, 25 Caucasians, 5 Blacks, and 4 Hispanics.

The population PK model was developed as a consequence of PK knowledge discovery performed on the data described above. Briefly, graphical displays were used for structure revelation and hidden patterns in the data. Thereafter, one- and two-compartment PK models with first-order input were tested for their ability to appropriately characterize the pharmacokinetics of the drug using the NONMEM software. The PK data were best described with the two-compartment model incorporating a first-order input. The parameters of the model were: absorption rate constant, apparent volume of the central compartment, apparent volume of the peripheral compartment, intercompartmental clearance, and apparent clearance. Empiric individual Bayesian post hoc parameter estimates were obtained and subjected to more exploratory data analysis (i.e., graphical analysis and generalized additive modeling—GAM) for initial covariate selection. The GAM analysis was coupled with bootstrap replication stability (13) to select covariates with inclusion frequency of ≥50% from 100 nonparametric bootstrap replicates. The covariates—age and dose level—selected by GAM were used to create a full model in NONMEM from which an irreducible final model was obtained by backward elimination. The irreducible model for the key parameter of interest—apparent clearance—included age and dose level as significant predictors. Thus,

$$CL/F = 30.2 - 3.22 \cdot (\text{Age} - \text{Median}) * IND + 0.14DL$$

where IND takes on the value of 0 if Age is greater than the median age and 1 otherwise, and DL is dose level.

The model was used to simulate concentrations from which area under the plasma concentration–time curve (AUC) was computed and compared with AUC calculated from model-based parameters as a means of checking the posterior predictive performance of the model. Figure 32.1 shows the distribution of AUC values for observed (model-based) and simulated data for the two studied doses and the 100 mg target dose. The results of 10 replications out of 100 are shown in the figure for illustrative purposes. The median simulated AUC for the 600 mg dose was similar to the model calculated AUC and over 88% of the AUC values from the simulated data overlap the model calculated AUC. The median AUC values obtained from the simulated data for the 200 mg dose were slightly biased (18%) with 68% of the values overlapping the model calculated AUC values. The Kolmogorov–Smirnov goodness-of-fit test, a two-sample comparison test, was used to perform the posterior predictive check for each replicate. The null hypothesis was that the observed (model based) AUC distribution and the simulated ones were equal, with the alternative that they were not. The ranges of p-value obtained across replicates for the 200 and 600 mg doses were 0.11 to 0.75 and 0.24 to 1, respectively, indicating similarity in the distributions. Although the model performed better in predicting AUC with the 600 mg dose than with the 200 mg dose, the population PK model developed did provide a reasonably adequate description of the data and was later employed to simulate PK profiles for virtual subjects used in data supplementation described below.

With PK knowledge discovery performed, the PK/PD knowledge creation was then performed in two phases as follows.

Phase I consisted of three steps:

FIGURE 32.1 Distributions of AUC values for 200 and 600 mg dose groups studied with parallel comparisons of those obtained via posterior predictive performance check (adapted from Ref. 2). AUC values from PK data simulated for the target dose of 100 mg are also included for comparison. "Observed" in the figure refers to population PK model-based AUC. See Appendix 32.1 for the S-Plus code used to generate the plot.

Step 1. Simulation of subjects with demographics similar to those in real study data set. Briefly, data synthesis of covariates for the virtual study was done through a resampling with replacement approach to ensure that the covariate distributions in each virtual study were similar to the real study. The demographic data from the real study was examined for correlation between covariates. There were no significant correlations between age and gender, and age and race. Also, gender and race were not correlated. However, age and weight were correlated. Given the total number of subjects, n, in the data set, a covariate vector such as gender was resampled with replacement from the observed data so that the proportion of males and females in the simulated data set was similar to that in the real data. This procedure was repeated for the other uncorrelated covariates. Where the covariates were correlated the resampling was done at the subject level to maintain the correlation structure. This resampling with replacement approach ensures equal probabilities for each element (covariate) of the population. The above algorithm was replicated for each virtual study used in the multiple data supplementation step discussed later. Figure 32.2 illustrates the distributions of some of the covariates. It is worthwhile to note the similarity between the distributions of the resampled covariates and those from the real study.

Step 2. Simulation of PK profiles for subjects from step 1 using a population PK model developed from data obtained from previously completed trial, and computation of exposure metrics and PK parameters for the simulated subjects. The distribution of AUC values, for instance, the target dose of 100 mg, is shown in Figure 32.1. These AUC values obtained from the first 10 replications were used for data supplementation.

FIGURE 32.2 Comparison of distributions of demographic variables between the real and simulated data sets (adapted from Ref. 2). M1 to M10 represent the number of replications used for data supplementation. See Appendix 32.2 for a sample S-Plus code.

Step 3. Combination of individual PK (exposure) variables from virtual subjects with subject-specific covariate data together with the real data set (including the pharmacodynamics–biomarker response data) to create a PK/PD knowledge creation data set.

Phase II was performed in two major steps:

Step 1. Performance of data structure analysis on real study data (untransformed and transformed) to reveal hidden structure, patterns, and relationships in the data set. This involves data visualization (graphing and fitting) and exploratory modeling (e.g., tree-based modeling).

Step 2. MS is used to generate ($M = 10$) baseline biomarker values for simulated subjects in the target dose group in which knowledge is to be created. MS is performed based on the revealed data structure. Figure 32.3 provides the schema of an example of the SBMS approach. In the example under consideration, the target dose group was partitioned into two groups: group A—subjects with higher biomarker baseline values and younger age (likely responder group)—and group B—the remainder of the subjects (the likely nonresponder group). Reduction of biomarker levels from baseline value was an indication of subject response to therapy. Basically the data supplementation for group A proceeded as follows: (a) the slope from day 0 to day 8 was supplemented from the real data that contained subjects matching the subpopulation criteria from day 0 to day 8; and (b) slopes for other time periods (i.e., day 8 to 15 and day 15 to day 28) were supplemented from all available data at the same time periods.

FIGURE 32.3 Schematic of the structure-based multiple supplementation (SBMS) approach (adapted from Ref. 2).

For group B subjects, biomarker responses were supplemented from all available PD data to reflect the overall uncertainty. Altogether, 10 replicates of target group data sets were created for the target dose group. The S-Plus code for the implementation of structure-based multiple supplementation is in the appendix.

Figure 32.4 shows the transformed data distributions of percentage changes from baseline values (i.e., slope) across replicates for the 100 mg target dose group from day 0 to day 8. In Figure 32.4A all subjects are included, but Figure 32.4B contains the responder subpopulation only. It can be observed that subjects who met the responder criteria (Figure 32.4B) had steeper slope values; the majority were in the −0.5 to −1 range.

After the creation of the biomarker data, each of 10 replicate data sets for the target 100 mg dose group was subjected to modeling and the results combined for PK/PD knowledge creation on the performance of the 100 mg dose level. The details of the modeling and results thereof are beyond the scope of this chapter. However, Figure 32.5 presents the data created for the target dose group in addition to the real data that were collected for the other groups that were studied. The results were consistent with the pharmacology of the drug. The supplemented biomarker response for the 100 mg target dose group appeared better than that observed for the 600 mg dose in the responder population as revealed by the slope, but comparable with the 200 mg dose (Figure 32.5). The knowledge created about the performance of the target 100 mg dose was communicated to the development team for the design of a future trial.

FIGURE 32.4 Day 0 to 8 slope distributions across multiple supplementation replicates for the target dose level: (A) all subjects and (B) subjects in the responder subgroup. (This figure is from Ref. 2.)

32.6 IMPLICATIONS OF THE USE OF MULTIPLE DATA SUPPLEMENTATION FOR THE CHARACTERIZATION OF AN UNEXPLORED REGION OF THE RESPONSE

The success rate of new chemical entities (NCEs) is anything but stellar (14). In 1987 the cost of bringing a new drug into the market was $237 million as opposed to $802 million in 2000 (15). By the end of 1999, 21% of the NCEs with investigational new drug applications (INDs) filed from 1981 to 1992 had been approved for market-

FIGURE 32.5   Slope distributions across days among different dose groups for subjects in the responder subgroup (adapted from Ref. 2). Note that 100 mg is the targeted dose group.

ing in the United States (16). Of those that failed in the period from 1987 to 1992, 38% of the NCEs failed because of efficacy (activity too weak or lack of efficacy), 34% because of economics (commercial market too limited, or insufficient return on investment), 20% because of safety (human or animal toxicity), and the rest for nonspecific reasons (16). What is becoming increasingly clear is that traditional drug development approaches are unlikely to succeed in the future given the economics of drug development—a low probability of success coupled with increasing product development times means decreased sales time after market launch and lower return on investment for pharmaceutical companies.

To speed drug development, sophisticated new technologies and approaches in the discovery and design of new drugs are replacing the traditional methods of discovery. Increasingly, however, a pharmacometrically guided approach is being applied to drug development. The need to get the most knowledge from every drug development study that is performed cannot be overemphasized in this day and age of spiraling drug development cost. This need has led to the development of PK/PD knowledge creation in general, and data supplementation in particular. In the application example, the nature of the response to a targeted dose was obtained. The drug effect that would have been produced if the target dose of 100 mg was studied would have been better than that produced by the 600 mg dose, but comparable with that produced by the 200 mg in the responder group. Thus, the drug was postulated to have an inverted U-shaped dose–response curve with doses above 200 mg producing lesser effect than the 200 mg dose. The application of PK/PD knowledge creation and implicitly knowledge discovery during drug development will optimize the drug development process and promote rational pharmacotherapy. Concerning drug development, Minto and Schneider (17) have stated: "Rapidly evolving changes in health care economics and consumer expectation make it unlikely that traditional

drug development approaches will succeed in the future. A shift away from the narrow focus of rejecting the null hypothesis toward a broader focus of seeking to understand the factors that influence the dose–response relationship together with the development of the next generation of software based on population models should permit a more efficient and rational drug development programme." The drug development process can be improved by implementing knowledge-driven development strategies founded on powerful, informative, and robust clinical trials. PK/PD knowledge creation and its companion knowledge discovery processes play pivotal roles in the generation and extension of knowledge and therefore can be influential in bringing efficiencies to the drug development process.

PK/PD knowledge creation via data supplementation results in the further acquisition of knowledge beyond that embedded in one's data. When data supplementation for PK/PD knowledge creation is implemented, the result is a greatly improved understanding of the response surface because of the knowledge created. This in turn leads to efficient design studies.

The approach to data supplementation described in this chapter can only be implemented if the conditions stipulated in Section 32.4.1 Motivation for Data Supplementation are met. The MS approach and its modification, SBMS, are an adaptation of the MI paradigm with the focus being data supplementation and not "missingness." From Figure 32.4 it is obvious that at least five replications are sufficient for obtaining robust data from the data supplementation process. MS should not be confused with missingness and the conditions that must be satisfied before performing MS. With MS, the focus is on characterizing an unexplored region of the response surface that was not the subject of focus in an already completed study. With the MS approach, data are generated for knowledge creation or acquisition, while MI deals with filling in missing data to create a complete data set for analysis.

It must be stated that in the current climate insufficient attention is given to knowledge-based drug development. The process of drug development can be no better than the knowledge on which it is based. Without adequate knowledge it is impossible to have a thorough understanding of one's drug, with the consequent compromising of the optimal development strategy. PK/PD knowledge creation through data supplementation, with knowledge discovery being an implicit component of it, results in gaining knowledge about a targeted region of a response surface that was not previously studied in a completed study without expending resources in conducting a new study. This provides the drug developer with the wisdom to make the right decisions about future trials and the strategic path for the development of a drug.

REFERENCES

1. P. J. Williams, C. J. Godfrey, A. Roy, H.-M. Chu, and E. I. Ette, Pharmacokinetics/pharmacodynamics knowledge discovery and creation during drug development, in *Industrial Pharmacokinetics*, Vol. 2, P. Bonate and D. Howard (Eds.). AAPS Press, Arlington, VA, 2004, pp. 163–202.
2. E. I. Ette and H.-M. Chu, Data supplementation: a pharmacokinetic/pharmacodynamic knowledge creation approach for characterizing an unexplored region of the response surface. *Pharm Res* **22**:523–531 (2005).

3. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**(3):275–291 (1997).

4. E. I. Ette, P. Williams, H. Sun, E. Fadiran, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).

5. E. I. Ette and L. C. Onyiah, Estimating inestimable standard errors: the bootstrap and winsorization. *Eur J Drug Metab Pharmacokinet* **27**(3):213–224 (2002).

6. D. R. Jones, Computer simulation as a tool for clinical trial design. *Int J Biomed Comput* **10**:145–150 (1979).

7. K. L. Lee, M. Frederick, C. F. Starmer, P. J. Harris, and R. A. Rosati, Clinical judgment and statistics: lessons from a simulated randomized trial in coronary artery disease. *Circulation* **61**(3):508–515 (1980).

8. D. B. Rubin, *Multiple Imputation for Nonresponse Surveys*. Wiley, Hoboken, NJ, 1987.

9. D. B. Rubin, Inference and missing data. *Biometrika* **63**:581–582 (1976).

10. J. L. Schafer, *Analysis of Incomplete Multivariate Data*. 2nd ed. Chapman & Hall/CRC, Boca Raton, FL, 2000.

11. D. B. Rubin, The Bayesian bootstrap. *Ann Stat* **9**:130–134 (1981).

12. D. B. Rubin and N. Schenker, Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc* **81**:366–374 (1986).

13. E. I. Ette, Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* **37**:486–495 (1997).

14. M. I. Kleinberg and L. A. Wanke, New approaches and technologies in drug design and discovery. *Am J Health Syst Pharm* **52**:1323–1336; 1341–1343 (1995).

15. C. Connolly, Price tag for a new drug: $802 million; findings of Tufts University study are disputed by several watchdog groups. *Washington Post*, 1 December 2001.

16. J. A. DiMasi, Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **69**:297–307 (2001).

17. C. Minto and T. Schneider, Expanding clinical applications of population pharmaco-dynamic modeling. *Br J Clin Pharmacol* **46**:321–333 (1998).

APPENDIX 32.1   CODE FOR AUC SIMULATION BASED ON POPULATION PK MODEL

```
################################################################
# to simulate the AUC using Pop. PK model ( Clearance adjusted by age)
# tmp is the data object that contains 4 columns (ID, AGE, DLVL, AUC)
# DLVL is the dose level
# myAUC.ans2 is the observed data object
# myAUC.ans3 is the simulated data object
#################################

tmp_real.bm.sim3[real.bm.sim3$DAY=="Day 1-8hr",]
myAUC.ans2_tmp[,c("ID","AGE","DLVL","AUC")]
myAUC.ans2$Dose_myAUC.ans2$DLVL

mySD_.15 # mySD is CV (15% of CV)
n_200 # 200 virtual studies
for(i in 1:n)
```

```
{
    # x is to resample age for the each virtual study
    # x1 is the indicate variable (1 if age LE 45, otherwise 0 )
    x_sample(myAUC.ans2$AGE,nrow(myAUC.ans2))
    x1_rep(1,nrow(myAUC.ans2))
    x1[x>45]_0
    x2_rnorm(n=nrow(myAUC.ans2),mean=0,sd=mySD)        # 15% CV
    y_(15.1-1.61*x1+0.0724*myAUC.ans2$Dose)*exp(x2)   # y is the
    clearance
    y1_myAUC.ans2$Dose/y
    xz_data.frame(age=x,dose=myAUC.ans2$Dose,AUC=y1,rep=rep
    (i,length(x)))
    if(i ==1){myAUC.ans3_xz}
        else{myAUC.ans3_rbind(myAUC.ans3,xz)}
}

############################################################
# Code for Boxplot of AUC values
############################################################

frame()
x_myAUC.ans[myAUC.ans$DLVL>0,]

x1_boxplot(split(x$AUC,x$DLVL),plot=F)
x1$names_paste(x1$names,"mg observed")
x2_bxp(x1,width=c(0.2,0.2),boxwex=0.2)

# to add sim.copies of AUC by dose
x.pos_c(0,x2[1]+10,x2[2]+10)
x_myAUC.ans
x1_boxplot(split(x$AUC.sim1,x$Dose),plot=F)
boxes(x=x.pos,width=c(1.2,1.2,1.2),stats=x1$stats)

y_seq(9,25,by=2)

for(i in 1:length(y))
{
    x.pos_x.pos+2
    xz_x[,c(5,y[i])]
    x1_boxplot(split(xz[,2],xz[,1]),plot=F)
    if(i==4)
    {
    boxes(x=x.pos,width=c(1.2,1.2,1.2),stats=x1$stats,labels=paste(
x1$names, "mg","\n","simulated"))
    }
    else{boxes(x=x.pos,width=c(1.2,1.2,1.2),stats=x1$stats)}
}

mtext(side=2,text="AUC (ug*hr/mL)",line=3,cex=1.2)
```

APPENDIX 32.2   CODE FOR DENSITY PLOT—OBSERVED VERSUS SIMULATED

```
####################################################
## to plot density plot for observed vs. simulations

par(mfrow=c(2,1))

myDose_200

plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose]),ylim=c(0,.6)
,xlim=c(0,20),lty=1,type="n",xlab="",ylab="Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=40,width="
nrd"),lty=1, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=10,width=
"nrd"),lty=2, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=20,width=
"nrd"),lty=3, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=30,width=
"nrd"),lty=4, lwd=4)

for(i in 1:n)
{
    lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose  &  myAUC.
ans3$rep==i]),lty=2,lwd=1)
}

key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(a) 200 mg",cex=1.2,line=1)

myDose_600

plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose]),ylim=c(0,.6)
,xlim=c(0,20),lty=1,type="n",xlab="",ylab="Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=40,width=
"nrd"),lty=1, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=10,width=
"nrd"),lty=2, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=20,width=
"nrd"),lty=3, lwd=4)
#lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose],n=30,width=
"nrd"),lty=4, lwd=4)

for(i in 1:n)
{
    lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose  &  myAUC.
ans3$rep==i]),lty=2,lwd=1)
}
```

```
key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(b) 600 mg",cex=1.2,line=1)


#######################################################
## to plot density plot for observed vs. simulations
## divided by age
#######################################################

par(mfrow=c(2,2))


n_50
myDose_100


plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE<45.5]),ylim=c(0,.6),xlim=c(0,20),lty=1,type="n",xlab="",ylab="
Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE<45.5],n=40,width="nrd"),lty=1, lwd=4)

for(i in 1:n)
#for(i in 1:10)
{
    lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose & myAUC.
ans3$rep==i]),lty=2,lwd=1)
}


key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(a) 200 mg of younger group)",cex=1.2,line=1)

plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE>45]),ylim=c(0,.6),xlim=c(0,20),lty=1,type="n",xlab="",ylab="
Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE>45],n=40,width="nrd"),lty=1, lwd=4)

for(i in 1:n)
{
    lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose & myAUC.
ans3$rep==i]),lty=2,lwd=1)
}
```

```
key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(a) 200 mg of older group)",cex=1.2,line=1)

myDose_600

plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE<45.5]),ylim=c(0,.6),xlim=c(0,20),lty=1,type="n",xlab="",ylab="
Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE<45.5],n=40,width="nrd"),lty=1, lwd=4)


for(i in 1:n)
{
    lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose  &  myAUC.
ans3$rep==i]),lty=2,lwd=1)
}

key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(a) 600 mg of younger group)",cex=1.2,line=1)

plot(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose  &  myAUC.ans2$
AGE>45]),ylim=c(0,.6),xlim=c(0,20),lty=1,type="n",xlab="",ylab="
Probability")
lines(density(myAUC.ans2$AUC[myAUC.ans2$Dose==myDose & myAUC.ans2$
AGE>45],n=40,width="nrd"),lty=1, lwd=4)

for(i in 1:n)
{
      lines(density(myAUC.ans3$AUC[myAUC.ans3$dose==myDose & myAUC.
ans3$rep==i]),lty=2,lwd=1)
}

key(
  text = list(c("Observed","Simulated"),cex=0.7),
  line = list(lty = 1:2, lwd = 4) ,corner=c(1,1)
)
mtext(side=3,text="(a) 600 mg of older group)",cex=1.2,line=1)
```

APPENDIX 32.3   SAMPLE CODE FOR PD DATA SUPPLEMENTATION

```
######################################################
# to simulate PD baseline values
# sim.data is the simulated data (a flat file) contains n(subjects)*
k(sample time points)
# rows with fields (columns, including demographics, exposure, dosing
regimen, etc)
# xy is the data set contains baseline,slopes and demo info from
the real data
# m copies of simulation
#sim.data.PD is the simulated PD data set (each column is one
simulation)
# impute=supplement

sim.data$PD_rep(NA,nrow(sim.data))
ID_unique(sim.data$ID)

m_10
n_length(ID)
sim.data.PD_sim.data[,c("ID","DAY1")]

myColumnName_c("ID","DAY2")
for(i in 1 :m)
{
sim.data.PD[,2+i]_rep(NA,nrow(sim.data.PD))
myColumnName_c(myColumnnName,paste("copy",i,sep=""))
}
dimnames(sim.data.PD)[[2]]_myColumnName
rm(myColumnName)
sim.data.PD$DAY2[sim.data.PD$DAY1 == 29] <- 28
sim.data.slope_sim.data.PD
t_matrix(rep(NA,n*m),nrow=n,ncol=m)
par(mfrow=c(2,5))
for(i in 1 :m)
{
   t[,i]_as.integer(runif(n,min=1,max=nrow(xy)))
   hist(t[,i],xlab="",ylab="")

}

# to impute baseline values based on t matrix
for(i in 1 :length(ID))
{
   for (j in 1 :m)
   {
      sim.data.PD[sim.data.PD$DAY2==0,2+j]_xy$PD[t[,j]]
   }

}
```

```
# to impute days 8, 15, 28

mySlope_xy$day8slope[xy$log.PD>=myTrim[3] & xy$AGE<=44.5 & xy$DLVL
<600]
mySlope_mySlope[!is.na(mySlope)]
mySlope2_xy$day8slope[xy$log.PD<myTrim[3] | xy$AGE>44.5 & xy$DLVL
<600]
mySlope2_mySlope2[!is.na(mySlope2)]
mySlope3_xy$day15slope
mySlope3_my$Slope3[!is.na(mySlope3)]
mySlope4_xy$day28slope
mySlope4_mySlope4[!is.na(mySlope4)]

# to impute based on the present criteria
for(i in 1 :n)


{
    age_sim.data$AGE[sim.data$ID==ID[i] & sim.data$DAY=="Predose"]
       for(j in 1 :m)
    {
       myBaseline_sim.data.PD[sim.data.PD$DAY2==0 & sim.data. PD$ID==
    ID[i],2+j]
    if (age<=44.5 & log10(myBaseline)>= myTrim[3])
    {
         t1_as.integer(runif(1,min=1,max=length(mySlope+.5)))
         myValue_myBaseline*(1+mySlope[t1])
         sim.data.slope[sim.data.slope$DAY2==8 & sim.data.slope$ID
==ID[i],2+j]_mySlope[t1]
    }
    else
       {
         t1_as.integer(runif(1,min=1,max=length(mySlope2+.5)))
         myValue_myBaseline*(1+mySlope2[t1])
         sim.data.slope[sim.data.slope$DAY2==8 & sim.data.slope$ID
==ID[i],2+j]_mySlope2[t1]

       }


    sim.data.PD[sim.data.PD$DAY2==8  &  sim.data.PD$ID==ID[i],2+j
    ]_myValue
    t1_as.integer(runif(1,min=1,max=length(mySlope3+.5)))
    myValue_myValue*(1+mySlope3[t1])
    sim.data.PD[sim.data.PD$DAY2==15  &  sim.data.PD$ID==ID[i],2+j
    ]_myValue
    sim.data.slope[sim.data.slope$DAY2==15  &  sim.data.slope$ID==
ID[i],2+j]_mySlope3[t1]
```

```
   t1_as.integer(runif(1,min=1,max=length(mySlope4+.5)))
   myValue_myValue*(1+mySlope4[t1])
   sim.data.PD[sim.data.PD$DAY2==28    &    sim.data.PD$ID==ID[i],
2+j]_myValue
      sim.data.slope[sim.data.slope$DAY2==28 &
   sim.data.slope$ID==ID[i],2+j]_mySlope4[t1]


   }
}
```

Clinical Trial Simulation: Theory

PETER L. BONATE

33.1 INTRODUCTION

The use of modeling and simulation (M&S) in the pharmaceutical industry can have practical and fiscal consequences. The 2004 estimate to bring a new molecular entity to market is over $800 million (1). Analysts at International Business Consulting Services (2) suggested that the pharmaceutical industry cannot continue to develop drugs as they currently are doing. Drug development must be redesigned to remain competitive. Those industries, like aerospace and chemical companies, that routinely use M&S save millions of dollars each year by doing so (3). To this end, the analysts at International Business Consulting recommended that in silico methods might be used to design better trials and, in so doing, possibly reduce the number of clinical trials that are performed, thereby reducing costs. Another consulting group, PricewaterhouseCooper, concluded that "virtual [clinical] trials will cut the amount of clinical resources required [to bring a drug to market] by 10% at best in the short term (although the savings may be much greater later on)," which translates into a savings of approximately $80 million initially. Hence, there is a strong financial incentive for companies to implement in silico approaches to drug development.

Clinical trial simulation (CTS) is the simulation of clinical trials within a computer and can be used to design powerful, robust, informative, and efficient clinical trials. Understanding CTS begins by understanding the theory behind simulation in general because CTS is simply a special type of simulation. CTS first gained notice through the work of Hale (4) by performing simulation as a prelude to running a registration clinical trial for mycophenolate mofetil—a transplant drug. Subsequently, CTS software programs were developed by MGA Software (Concord, MA) and Pharsight Corporation (Mountain View, CA) in the late 1990s. ACSL Biomed, MGA Software's product, was later acquired by Pharsight Corp. and its features were incorporated into later versions of Pharsight's Trial Designer product. Prior to the introduction of these products, simulation was limited to largely

deterministic models and was not implemented to any large degree. Today, large pharmaceutical companies, and even progressive smaller companies, have entire groups dedicated to the development of M&S of pharmaceutical data.

The purpose of this chapter is to introduce the concepts of simulation and use these concepts to illustrate how more complex pharmacokinetic pharmacodynamic (PK/PD) simulations may be designed. Furthermore, the chapter discusses recent uses of M&S by pharmaceutical companies and by regulatory authorities as examples.

33.2   MODELING VERSUS SIMULATION

A system is an interacting or interdependent group that forms a whole or a set of interrelated components. Ludwig von Bertalanffy (1901–1972), a notable theoretical biologist who studied systems biology, once said that "oversimplifications, progressively corrected in subsequent development, are the most potent or indeed only means toward conceptual mastery of nature." Modeling is one such means to simplify a system. Modeling reduces a complex system, one that cannot be easily understood on its own, into a simpler representation. For instance, a scale model of a building is an architectural model. In pharmacometrics the emphasis is on mathematical and statistical models of systems. As models develop, greater and greater understanding of the system is achieved.

Two types of mathematical models are encountered: empirical and mechanistic (sometimes called theoretical). Empirical models are data-driven, black-box models that simply act to describe the data. An example of an empirical model is a time-series analysis or a compartmental fit to a PK profile. Increasingly, however, models are becoming mechanistic, incorporating what is known about the system into the model (see Refs. 5–7 for recent examples). For example, transport to tissues, blood flow, saturable protein binding, or in vitro enzyme kinetic parameters may be built into the model a priori. Currently, a number of companies are developing commercial mechanistic models of disease states like asthma, obesity, and acquired immune deficiency complex (Entelos, `www.entelos.com`), models of tissues or organs (Physiome, `www.physiome.com`), or models of oral drug absorption (Gastroplus, `www.simulationsplus.com`).

There are advantages and disadvantages to both model types. Empirical models require fewer assumptions, are easy to develop, and are useful for prediction. Mechanistic models require more assumptions, may not be easy to develop, are useful for prediction, and, importantly, can be used for extrapolation. Many types of empirical models are limited in their ability to extrapolate outside the bounds of the data-generating mechanism. Truly complex models may consist of both mechanistic and empirical domains with some parts of the model being mechanistic and other, unknown parts of the system being empirical.

The type of model that one chooses to use depends on the reason for developing the model. If the reason is to characterize a set of data into a few summary parameters, such as reducing a concentration–time profile into an individual's clearance, volume of distribution, and half-life, then an empirical model will likely be chosen. On the other hand, if the goal is to understand how a patient's exposure might change if renal function is impaired, then a more mechanistic model will likely be

chosen. The key here is to choose a model that will answer the question efficiently and accurately.

Empirical and mechanistic models are examples of mathematical models. Given a set of observed data from a known input, often the goal is to identify a model and estimate its associated parameters—the problem of inverse estimation. In order to reach this goal, however, it is not enough to have a mathematical model; a statistical model is also needed. Statistical models describe the distribution and sources of variability in the observed data. For example, it may be assumed the only source of error at an individual level is random error from the measurement of the data. Alternatively, it may be assumed that error varies not only within an individual on a given day, but may vary across days as well, so-called interoccasion variability. On a population level where the analysis focuses on data from many individuals, the sources of variability may be between-subject variability, interoccasion variability, and residual variability. Once the sources of variability are identified, their probability density function must be defined. A common assumption in modeling is to assume that random error is normally distributed. In pharmacokinetics, it is often assumed that between-subject variability is lognormal, as are residual variability in concentration data. Given a mathematical and statistical model, the problem of parameter estimation may be solved through either Fisher (least-squares or some modification thereof) or Bayesian methods, but usually the former.

Prior to model estimation the question that it will be used to answer and the specific manner in which it will be used should be explicitly stated. Using a model to answer a question is the act of simulation. There are two types of simulation: deterministic and stochastic. In a deterministic simulation, the statistical model is ignored and no error is introduced into the model—the results are error-free. For example, given data from single-dose administration of a drug it may be of interest to predict the typical concentration–time profile at steady-state under a repeated dose administration regimen. A deterministic simulation would be useful in this case.

Deterministic models are of limited value. In a stochastic simulation, error is introduced into the model as a random draw from the statistical probability distribution, that is, the statistical model. Stochastic simulation is often referred to by its more common moniker—Monte Carlo simulation, a term invented by von Neumann and Ulam, two scientists working on the Manhattan Project during World War II, who coined the term because of its association with gambling casinos in Monte Carlo (8). Because of the random nature of the data, many different outcomes are possible and because many different outcomes are possible, a stochastic simulation frequently examines the long-term nature of an outcome or the extremes of an outcome. As an example, suppose that there is an 80% chance of a patient experiencing severe nausea and vomiting when average steady-state concentrations exceed some value. Assuming that the population distribution for clearance is known, for example, lognormal with a mean of 15 L/h and a 30% coefficient of variability (CV), and given a fixed dose regimen, then one could use stochastic simulation to estimate the percentage of patients who will exceed the cut-off concentration value. If a large percentage of patients exceed the adverse event cut-off value, then a dose reduction may be indicated.

Bayesians add an additional layer of error to a simulation. Fisherians treat the estimated parameters obtained from fitting a model as fixed quantities, when clearly

they are not. For example, in the previous example, the mean clearance was fixed to 15 L/h and the CV was fixed to 30%. These estimates can never be known with certainty; they were estimated somehow. And because they are estimates there is some measure of imprecision in their values. Bayesians take this additional level of uncertainty into account when performing a simulation. For example, the mean clearance may be estimated as 15 L/h with a standard error (called the standard error of the mean) of 1 L/h. Because of the central limit theory, the distribution of the mean is assumed to be normally distributed. The distribution of a variance cannot be normally distributed because variance components can never be negative. Hence, it is common to assume that the inverse of variance components have a gamma (univariate case) or Wishart (multivariate case) distribution (9). So, for instance, instead of fixing clearance from a lognormal distribution with mean μ and variance σ^2, which is how a Fisherian would simulate from a model, a Bayesian includes a hyperparameter step whereby μ has its own probability distribution, as does σ^2. Therefore, before drawing a random variate from $X \sim (\mu, \sigma^2)$ two other random draws are made first. One being from $\mu \sim (\mu_\mu, \sigma_\mu^2)$ and the other from $1/\sigma^2 \sim \text{Gamma}(\alpha, \beta)$, which are the hyperparameters for the model parameters. For example, in a Bayesian simulation, μ may take on values 14.7, 15.2, and 15.8 L/h and CV may take on values 28.9%, 35.0%, and 40.0% in the first three iterations of the simulation. μ may converge to the value of μ used in the Fisherian simulations over the long run, but not necessarily. Frequently, Bayesian simulations lead to wider confidence intervals than Frequentist simulations because of the extra layer of randomness built into the simulation (see Figure 33.1).

Hence, there are subtle but important differences between modeling and simulation. Modeling looks back in time, whereas simulation looks forward. Simulation is used to answer "What if . . . ," whereas modeling is used to answer "What happened?" Modeling requires data, simulation requires models that may or may not be built on data. Both M&S are sensitive to their assumptions and any blackbox aspects in their use, both of which may lead to criticism of a M&S exercise. Furthermore, both M&S can be used to identify important variables and can be used to summarize and understand complex systems.

33.3 ELEMENTS OF SIMULATION

Complex pharmacokinetic/pharmacodynamic (PK/PD) simulations are usually developed in a modular manner. Each component or subsystem of the overall simulation is developed one-by-one and then each component is linked to run in a continuous manner (see Figure 33.2). Simulation of clinical trials consists of a covariate model and input–output model coupled to a trial execution model (10). The covariate model defines patient-specific characteristics (e.g., age, weight, clearance, volume of distribution). The input–output model consists of all those elements that link the known inputs into the system (e.g., dose, dosing regimen, PK model, PK/PD model, covariate-PK/PD relationships, disease progression) to the outputs of the system (e.g., exposure, PD response, outcome, or survival). In a stochastic simulation, random error is introduced into the appropriate subsystems. For example, between-subject variability may be introduced among the PK parameters, like clearance. The outputs of the system are driven by the inputs

FIGURE 33.1   Box and whisker plot of average steady-state concentration for 500 simulated subjects as a function of dose using Fisherian (F) and Bayesian simulation (B). Clearance was assumed to be lognormally distributed with a median of 15 L/h and a CV of 30%. The parameters in the Fisherian simulation were fixed. In the Bayesian simulations, median clearance was assumed to be normally distributed with mean 15 L/h and a standard error of 1 L/h. $1/\sigma^2$ was assumed to have a gamma distribution with parameters (0.8, 15) leading CV to have a range of about 20–54% with a mean of 30%. This data could then be applied to a pharmacodynamic model, such as probability of experiencing nausea or vomiting. In this case, the solid line denotes the average steady-state concentration leading to a 20% probability of a patient experiencing severe nausea and vomiting. At doses up to 40 mg once daily, no subjects will reach the cut-off value, whereas a significant proportion will with a daily dose of 80 mg.

FIGURE 33.2   Flowchart of a PK/PD simulation at an individual level. Shaded boxes denote stochastic elements. Arrows denote the flow of information and inputs/outputs from a model component. Lines with solid circles denote sampling components.

into the system and may reflect something as simple as exposure (area under the curve or maximal concentration) or something more complicated, such as survival. To be realistic, outputs may be missing at random, censored, or incorrectly coded into the database. In a theoretical simulation, missing data are frequently ignored but in an actual CTS how missing data occur is very important and can affect the simulation's outcome.

Layered on the input–output model is the trial execution model, which defines how the clinical trial is conducted. It is at this stage that the dosing regimen is defined. For example, subjects are randomized to three doses: 10, 20, and 40 mg daily for 3 months. At baseline and each month thereafter, PD response is assessed. The dosing regimen may be adaptive in the sense that dose reductions are made if a patient reaches some exposure threshold that triggers an adverse event. Also defined at this stage are protocol deviations. Urquhart (11) defines three types of deviations: study initiation deviations, protocol compliance deviations, and termination deviations. Examples of study initiation deviations are when patients who do not meet the inclusion requirements are accidentally enrolled or patients might refuse to enroll in the study. An example of a protocol compliance deviation, which is the most common deviation in a study, is when patients do not intake drug with equal probability at the same time each day. Instead, dosing could be modeled as a process that may change over time and at different times of day. For example, dosing could be modeled using the following: at each dosing event a uniformly generated random number on the interval [0, 1] may be simulated and that subject's dose is not taken if the value is ≤0.01. Hence, there is a 1% chance of a subject forgetting to take a dose. Alternatively, dosing compliance could be modeled using a more complex Markov chain model (12). Another type of protocol compliance deviation is the patient who misses an office visit. Hence, data for that day are missing. Or the patient may take a drug that is contraindicated in the protocol. The last type of deviation is termination deviations, where individuals drop out from the study either randomly or for treatment-related reasons. Trial execution models should be as simple or complex as needed.

Once an input–output model is coupled to a trial execution model, the trial may be simulated either once or many times. Each time the simulation is executed is referred to as an iteration, replicate, or run. For troubleshooting purposes, one should always confirm that the simulation works using a few replicate runs, two to five should be sufficient, before progressing to executing the simulation many times. Many stochastic simulations may take days to run and if the simulation does not work as performed using a single run, then days may be wasted before finding this out. In a deterministic simulation, there is no need to replicate the simulation many times because the outcome will always be the same. However, with a stochastic simulation, the outcome could be different every time the simulation is replicated.

Sun et al. (13) differentiate two types of clinical trial simulations: computer-assisted trial designs (CATDs) and computer-simulated clinical trials (CSCTs). CATDs are used to help guide actual, future clinical trials through simulation, allowing a user to "test-drive" a clinical study design before actually doing the study. CSCTs can be used to determine clinical trial outcomes without ever having to do the study and are designed to prevent having to do such a trial. Sun et al. (13) gave as an example of CATD the label change for Augmentin® 500 mg three-times-daily to 875 mg two-times-daily. Simulations showed similar efficacy rates between the

two dosing regimens, which was then confirmed by one clinical trial. An example of a CSCT might be a pharmaceutical company that uses simulation as an argument to regulatory authorities that the change in exposure is negligible if a 50% decrease in hepatic clearance were to occur, thus little new information would be gained in doing a study in patients with hepatic impairment.

A key question in any Monte Carlo simulation is: How many replications are needed for suitable accuracy and precision? Often the number of replications is based on ad hoc rules. For example, usually 50 or more replicates are sufficient if the goal is to obtain an estimate of an average outcome. However, much larger numbers of replicates are needed if the goal is to observe rare events, numbers often in the thousands, because rare events rarely occur and large numbers are needed to see them. Large replicate numbers are also needed to obtain confidence intervals or estimates of variance components. These ad hoc rules are loosely based on the central limit theorem and Chebyshev's inequality. Assume $\hat{\theta}_i$, $i = 1, 2, \ldots, n$, are independent estimates of θ, the parameter of interest. Let $\bar{\theta}$ be the estimator for θ and let S be an estimate of the standard deviation of $\hat{\theta}$, also called the standard error of the estimate. The central limit theorem is based on the law of large numbers and roughly states that, for large n, $(\bar{\theta} - \theta)/(S/\sqrt{n})$ is approximately distributed as a standard normal random variable. So, for example, there is no more than a 5% chance that $\bar{\theta}$ is more than 1.96 standard errors from θ. Chebyshev's inequality, which is more conservative than the central limit theorem, states

$$\left(p\left(|\bar{\theta} - \theta| > c \frac{\sigma}{\sqrt{n}} \right) \leq \frac{1}{c^2} \right) \tag{33.1}$$

In other words, the probability that $\bar{\theta}$ is more than c standard errors from θ is less than or equal to $1/c^2$. So there is no more than a 26% chance that $\bar{\theta}$ is more than 1.96 standard errors from θ. In both cases, as n increases, the standard error decreases, and $\bar{\theta}$ converges in probability to θ.

Ad hoc rules, while useful, may not be rigorous enough to some modelers. Fortunately, more statistically stringent criteria are also available (14). If the goal is to estimate the expected value of some variable, that is, $E(\theta)$, then first generate 30 replications of the simulation. Let $\hat{\theta}_i$ be the ith estimator for θ, $i = 1, 2, \ldots, n$. Let S be the standard deviation for all $\hat{\theta}_i$ and let $SE(\theta)$ be the standard error of the mean estimator, $\bar{\theta}$, which is defined by the modeler a priori. Continue replicating until $S/\sqrt{n} \leq SE(\theta)$, making sure that S and n are updated after each replicate is completed. The estimate for θ is then given by $\bar{\theta} = (\Sigma\hat{\theta}_i)/n$. Another similar method is based on the $100(1 - \alpha)\%$ confidence interval for θ. First, choose an α-level (i.e., 0.05) and then replicate the simulation until $2Z_{\alpha/2} S/\sqrt{n} < l$, where l is the minimum acceptable length of the confidence interval defined by the modeler a priori, updating S and n at each replication. There is very little difference between the outcomes of these rules and either may be used.

For a stochastic simulation replicated many times, it would be inefficient to analyze every single trial one by one. Besides, in a stochastic simulation, the analyst is not interested in what happens with a single trial but what happens in the long run. Recall that the probability of observing a sample drawn from a continuous probability distribution is zero; that is, the probability of observing the number

1.65 from a standard normal distribution is zero, but the probability of observing a number at least as large as 1.65 is 0.05. The same analogy holds in simulation. The probability of observing the results from any particular replicate among many replicate simulations is zero. However, the probability of determining an aggregate of results from across many different simulations may be determined.

Hence, CTS cannot answer the question "Will this trial succeed?" but rather answers the question "What is the probability of this trial succeeding?" One outcome frequently reported from a CTS is that of statistical power—the probability of declaring the null hypothesis (the hypothesis of no difference) false. Power is computed from a stochastic simulation by counting the number of times the null hypothesis is rejected divided by the total number of replicates used in the simulation. For example, if 88 runs out of 1000 reject the null hypothesis, the power of the trial design is 8.8%. Since this type of analysis is binomial (the trial either fails or it succeeds), the $(1 - \alpha)\%$ confidence interval for the power of the trial may be calculated using the normal approximation as

$$\hat{p} \pm Z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \qquad (33.2)$$

where Z is from the standard normal distribution with $\alpha/2$ critical values and $\hat{p}$ is the estimated proportion of successes, 0.088 in this case. Hence, it may be concluded that under the assumptions made in the simulation, there is an 8.8% chance with 95% confidence interval of {7.0%, 10.6%} that the clinical trial detects a significant difference between treatments.

At this point, a sensitivity analysis may be undertaken to determine which factors have undue influence on the outcome of the simulation or have little effect. For instance, it may be found that in a trial testing doses of 10, 20, and 40 mg that no difference between the 20 and 40 mg dose is detected in 90% of the simulations. Hence, it may be recommended to remove the 20 mg dose from the study as it provides no additional information. Alternatively, parameters of the trial may be manipulated to increase the power of the simulation. For example, if dosing is changed to twice a day from once a day dosing, how does the power of the trial change and how does this affect cost to the patient?

33.4 RANDOM NUMBERS[1]

The backbone of Monte Carlo simulation is the ability to generate random numbers because random numbers form the basis of the random draws from a probability distribution. Computers, because they are based on rules, algorithms, and mathematical operations, cannot generate truly random numbers. Instead, random numbers start from some point in the algorithm, called the seed, and proceed in a linear, predictable manner but when examined in the short term appear to be random. It

[1]Parts of this section are reprinted with permission from P. Bonate, Random Number and Random Variate Generation. Copyright © 2001, by author.

should be stressed that every random number generator (RNG) has its deficiencies. Early RNGs had poor randomness properties. To make matters worse, after repeatedly calling a RNG, eventually the numbers will begin to repeat. The total number of random number calls that can be made before the numbers begin recycling is called the period. A few years ago, RNGs were limited by periods of ~2^{31}, which if the modeler was not careful would result in replications within a simulation that were not random, but were instead correlated with previous replications. Modern RNGs in use today (which can be found in software like Matlab (Mathworks, Boston, MA) or the Pharsight Trial Designer (Pharsight, Mountain View, CA)) have periods of more than 2^{100}, which is more than sufficient for any clinical trial simulation. Some software packages, such as SAS (SAS Institute, Cary, NC), while still using perfectly valid RNGs, have the old period limitations associated with them. But other software packages, particularly Microsoft Excel, not only have small periods (~2^{32}) but fail to generate random numbers that are even close to being random (15–17).

In order to understand how random numbers are generated, a brief diversion is needed in explaining modular arithmetic. Modular arithmetic is sometimes called clock arithmetic because it is similar in principle to how we tell time (i.e., it is arithmetic on a circle instead of a line). For instance, on a clock 7 + 7 is 2. In this case the period is 12h and the operation just described can be written

$$(7 + 7) \bmod 12$$

or

$$\bmod(7 + 7, 12)$$

where mod is called the modulus and is simply the remainder of the division of X/Y. In this case, 14/12 has remainder 2.

Most random numbers are generated using a multiplicative congruential algorithm (MCA) (18). If X_i is the current random number, then the next random number is

$$X_{i+1} = (aX_i)(\bmod M) \tag{33.3}$$

where a is a positive integer called the multiplier and M is the modulus. To obtain a number in the interval [0, 1), X_{i+1}/M is returned instead of X_{i+1}. Congruential algorithms must have a starting point called the seed, which is often an integer. The theoretical maximum period is given by M, but in reality, for MCAs the period will be much shorter. Probably the best known and widely used MCA is the generator by Park and Miller (19). The parameters for their generator are $M = 2^{31} - 1$ and $a = 16,807$. Under the algorithm the smallest numbers possible are 0.00000000046566 and 0.99999999953434. Notice that 0 and 1 are not included in the possible interval. This is a common characteristic of MCAs; 0 and 1 are not possible. This algorithm repeats itself after $2^{31} - 2$ numbers are generated. While this may seem a large number, on a Pentium class PC the algorithm can be exhausted in as little as a couple of hours.

A modification of the multiplicative generator is the linear congruential algorithm (LCA) or generator. If X_i is the current random number, then the next random number is

$$X_{i+1} = (aX_i + c)(\text{mod } M) \tag{33.4}$$

where c is a positive integer greater than zero called the constant. LCAs have better nonrepeating properties than MCAs. As a simple example, if $M = 16$, $a = 2$, $c = 3$, with a starting seed of 2, then the sequence of random numbers is shown in Table 33.1.

Once the sequence gets to 13, it repeats ad infinitum. This is clearly not a good RNG. M, a, and c must be chosen very carefully and this is where RNGs get into trouble. What distinguishes a good RNG from a bad RNG is the choice of a, c, and M. First, c and M can have no common divisor. Second, $a - 1$ should be a multiple of p, where p is every prime number that divides M. Third, if 4 divides M exactly then 4 must divide $a - 1$ exactly. In the example above, c and M have no common prime, which meets the first criterion. And there is only one prime number that divides 16 and that is 2. Therefore, $a - 1$ must be a multiple of 16 or 2, 4, 6, 8, 10, 12, or 14. Since 16 is divisible by 4, then $a - 1$ must be divisible by 4 as well. That leaves 4, 8, and 12. Then a can equal 5, 9, or 13. If $a = 9$ then the following sequence of numbers is generated: 2, 5, 0, 3, 14, 1, 12, 15, 10, 13, 8, 11, 6, 9, 4, 7, 2, ... Notice that the entire sequence is used before recycling. This is a trait of a good RNG. Two last things need to be pointed out here. First, with a LCA, it is possible for zero to be a generated value, although one may still not occur. Second, suppose that one sorted the possible values obtained using the generator $M = 16$, $a = 9$, $c = 3$, then the following random numbers are obtained: 0.00000, 0.06250, 0.12500, 0.18750, 0.25000, 0.31250, . . . , 0.81250, 0.87500, 0.93750. Notice that numbers between 0 and 0.0625 do not occur, nor do numbers between 0.0625 and 0.125, and so on. All random number generators produce discrete values and this discreteness is one reason why they are never truly random. Thus, RNGs are really pseudo-RNGs.

Some of the more current RNGs use alternative algorithms, like nonlinear recursion or multiple recursion, but these are beyond the scope of this chapter. These new algorithms have periods of more than 2^{100} (the Mercenne Twister has a period of $2^{19937} - 1$!) and therefore have better randomness properties. They are five- to tenfold slower than linear algorithms, however.

TABLE 33.1 The Linear Congruential Algorithm (LCA) Generator

X_i	$aX_i + c = 2X_i + 3$	Mod 16
2 (seed)	$2 \cdot 2 + 3 = 7$	7
7	$2 \cdot 7 + 3 = 17$	1
1	$2 \cdot 1 + 3 = 5$	5
5	$2 \cdot 5 + 3 = 13$	13
13	$2 \cdot 13 + 3 = 29$	13
13	$2 \cdot 13 + 3 = 29$	13
Repeats		

33.5 SIMULATING CONTINUOUS RANDOM VARIABLES

All random draws from a probability distribution begin with a "random" draw from a uniform distribution from the interval (0, 1) (see Figure 33.3). From uniformly generated random variates, most any probability distribution function (pdf) can be simulated using one of the following approaches:

- Inverse transformation
- Acceptance–rejection
- Convolution
- Decomposition (Composition)

Each of these methods is briefly discussed. The reader is referred to Ross (14) and Law and Kelton (20) for more details.

An example of inverse transformation is the exponential random variable X with mean $1/\lambda$ and variance $1/\lambda^2$, which has cumulative distribution function (CDF)

$$F(X) = 1 - \exp(-\lambda X), \quad X \geq 0, \lambda > 0 \tag{33.5}$$

FIGURE 33.3 Interrelationships between uniformly distributed random numbers and commonly used probability distributions. All probability distributions arise from the uniform distribution.

If we set U, a random variate on the interval $(0, 1)$, equal to $F(X)$ and solve for X, then

$$X = F(X)^{-1} = -\frac{1}{\lambda}\ln(1-U) \tag{33.6}$$

Hence, we can generate an exponential random variable by generating a random variable U and then use Eq. (33.6) to draw a sample from an exponential distribution with mean λ. This method of simulating continuous variables is called the inverse transformation method. Although the method can be applied to any distribution, either continuous or discrete, the problem with the inverse transformation approach is that it is often difficult to invert the CDF, if it even exists, to an analytical solution.

An example of the acceptance–rejection method is simulating from a Poisson distribution, which is frequently used to model count data. Poisson random variables have mean and variance λ. Poisson random variates can be simulated by the following algorithm:

Step 1. Generate a random number U on the interval $(0, 1)$.
Step 2. Set $m = 0$, $p = \exp(-\lambda)$, $F = p$.
Step 3. If $U < F$, set $X = m$ and stop.
Step 4. $p = \lambda p/(m + 1)$, $F = F + p$, $m = m + 1$.
Step 5. Go to step 3.

Another example of acceptance–rejection is an early method to simulate random draws from a standard normal distribution called the Box–Muller transformation (21). If U_1 and U_2 are random variates on the interval $(0, 1)$, then the variates

$$Z_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \tag{33.7}$$

$$Z_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \tag{33.8}$$

are independent standard normal deviates with mean 0 and variance 1. This method has been criticized because it is slow (it has many calls to functions within a math library (square root, cosine, and natural log)) and can be particularly unstable when U_1 is close to zero due to taking the natural log of a very small number. The advantage of the method is that there is a one-to-one transformation of uniform random variables to normally distributed random variables such that there is no loss of efficiency.

Once a standard normal deviate is generated, it can be transformed to a normal distribution with arbitrary mean μ and variance σ^2 by the transformation

$$X = \mu + Z\sigma \tag{33.9}$$

Often it is necessary to generate variates X and Y that are not independent but are correlated. In this case, bivariate correlated random variates can be generated as

follows. If Z_1 and Z_2 are independent standard normal deviates with mean 0 and variance 1, then let $X = Z_1$ and

$$Y = \rho Z_1 + \sqrt{1 - \rho^2}\, Z_2 \qquad (33.10)$$

where ρ is the desired correlation between X and Y.

A third method to simulate random variables is convolution, where the desired random variates are expressed as a sum of other random variables that can easily be simulated. For example, the Erlang distribution is a special case of the Gamma distribution when the shape parameter is an integer. In this case, an Erlang random variate with shape parameter β can be generated as the sum of β exponential random variates each with mean α. A last method to simulate random variables is decomposition (sometimes called composition), where a distribution that can be sampled from is composed or decomposed by adding or subtracting random draws into a distribution that cannot be simulated. Few distributions are simulated in this manner, however. These last two methods are often used when the first two methods cannot be used, such as if the inverse transformation does not exist.

33.6 SIMULATING DISCRETE RANDOM VARIABLES

Discrete random variables are actually quite simple to simulate. Suppose there are k discrete categories with probability p_i, $i = 1, 2, \ldots, k$, $\Sigma p_i = 1$. Determine the cumulative probabilities for each group. Then simulate a random variable U on the interval $(0, 1)$ and determine within which category U is found. For example, suppose smoking status was being simulated with three categories shown in Table 33.2. Now a random variate U was simulated with value 0.52. Since $0.52 \leq 0.67$, the corresponding category is "Not a smoker."

A similar algorithm can be applied for categorical variables with $k \times m$ categories. For example, suppose smoking status was then stratified by sex (to form a 3×2 table) such that the probabilities within each cell were as given in Table 33.3. This table can be broken down into the cumulative probability table shown as Table 33.4.

TABLE 33.2   Simulation of Smoking Status

Category	p_i	Cumulative Probability
Not a smoker	0.67	0.67
<1 pack a day	0.20	0.87
≥1 pack a day	0.13	1.00

TABLE 33.3   Smoking Status Stratified by Sex

Category	Males	Females
Not a smoker	0.32	0.35
≤1 pack a day	0.13	0.07
>1 pack a day	0.10	0.03

TABLE 33.4   A Cumulative Probability Table from Table 33.3

Category	Cumulative Probability
Not a smoker, male	0.32
≤1 pack a day, male	0.45
>1 pack a day, male	0.55
Not a smoker, female	0.90
≤1 pack a day, female	0.97
>1 pack a day, female	1.00

Now a random variate U with value 0.85 is simulated. Since $0.55 \leq 0.85 < 0.90$, the simulated value is a female who does not smoke.

33.7   SOFTWARE

Many different software packages exist for simulating clinical trials or PK/PD studies. Which one to use usually depends on the user. Some software is specifically designed to simulate PK/PD outcomes within the context of a clinical trial design, such as the Pharsight Trial Simulator. Then there are more general simulation languages with graphical user interfaces (GUIs), such as Simulink (Mathworks, Natick, MA). While GUIs are useful, they are often limited to what the developers include in the software, sometimes leaving the user caught between a rock and an upgrade. Programs such as NONMEM (Globomax LLC, Hanover, MD), WinNonlin (Pharsight Corp., Mountain View, CA), or ADAPT II (University of Southern California Biomedical Simulations Resource) were designed for model fitting and development, but can be used to simulate data. However, they must be used in conjunction with another program, like SAS (SAS Institute, Cary, NC) or S-Plus (Insightful Corp., Seattle, WA) as they cannot simulate data within the context of a clinical trial (i.e., they are dependent on the inputs supplied by the user and cannot simulate those inputs). Lastly, there are programming languages. While the most difficult to use, they are also the most flexible. These include Matlab (Mathworks, Natick, MA), Gauss (Aptech Systems, Maple Valley, WA), the IML procedure within SAS, and S-Plus. A common combination of software used to simulate clinical trials is the use of NONMEM and SAS or NONMEM and S-Plus, allowing the user to take advantage of NONMEM's large library of PK models (see Figure 33.4).

33.8   APPLICATION OF M&S IN DRUG DEVELOPMENT AND REGULATORY REVIEW

Much has already been written on the role of M&S in drug development (8, 10, 22–24). All of these reviews present examples of simulation in drug development. However, the examples reported in these reviews are often brief, lacking detail or insight. Rather than reprint what has already been reviewed, it was thought a more informative approach would be to present two case studies, one from the pharma-

FIGURE 33.4 Schematic illustrating how NONMEM and a more generalized software program, like SAS (SAS Institute, Cary, NC) or S-Plus (Insightful Corp., Seattle, WA), can be used to interact and simulate clinical trials.

ceutical industry and one from the regulatory point of view, in sufficient detail so as to illustrate the methodology.

Darbepoetin alfa (Aranesp®) is in the class of recombinant human erythropoietin proteins that has greater in vivo potency than recombinant human erythropoietin (r-HuEPO) through the addition of N-linked sialic acid side chains on the amino acid backbone of the protein. Darbepoetin alfa stimulates the production of red blood cells (RBCs) for the treatment of chemotherapy-induced anemia and in patients with chronic renal failure. The recommended dose is different between the indications. In patients with chemotherapy-induced anemia, the recommended starting dose is 2.25 μg/kg administered as a weekly subcutaneous (SC) injection with the weekly dose adjusted to maintain a target hemoglobin (Hgb). The dose should be increased to 4.5 μg/kg if the Hgb increase is less than 1.0 g/dL after 6 weeks of treatment. The dose should be reduced 25% if either the Hgb exceeds 12 g/dL or the increase in Hgb is more than 2.0 g/dL. Dosing should be withheld until the Hgb falls to at least 12.0 g/dL if the Hgb exceeds 13.0 g/dL, at which point dosing should be reinitiated at ~25% below the previous dose.

Alternative dosing strategies have been proposed to take advantage of the longer half-life of darbepoetin alfa compared to r-HuEPO. One regimen of 3 μg/kg every 2 weeks was shown to be equally efficacious as 40,000 U r-HuEPO once weekly. Dosing a patient based on their body weight may add an additional layer of complexity to the dosing regimen that might not be necessary. Hence, Jumbe et al. (25) used CTS to determine whether a fixed dose of darbepoetin alfa (200 μg every 2 weeks) has the same outcome as a weight-based dose of darbepoetin alfa (3 μg/kg every 2 weeks).

Data was pooled from three clinical trials (547 patients) studying the use of darbepoetin alfa in the treatment of chemotherapy-induced anemia. Serial PK and PD (Hgb concentrations) measurements were collected throughout the studies and merged into a single database along with patient-relative covariates. A population PK/PD model was developed that simultaneously modeled darbepoetin alfa

concentration–time profiles, as well as Hgb–time profiles (see Figure 33.5). The CTS was then developed encompassing the following elements:

- Model parameters and their associated between-subject variance estimates from the PK/PD model were fixed to their final values.
- The patient population demographics (body weight and baseline Hgb concentrations) were defined based on observed baseline values observed across the three studies.

The following study design elements were incorporated in the simulation:

- Dosing was every other week for 12 weeks by SC administration based on either weight or fixed dose.
- A transfusion was simulated if Hgb declined below 8.0 g/dL, whereby data from these patients were censored for the next 4 weeks.
- Dosing was withheld if Hgb was ≥14.0 g/dL in women or ≥15.0 g/dL in men.
- Censoring was randomly implemented to coincide with the censoring rates in the clinical trials.
- Other protocol elements, such as definition of response or sampling for PD analysis every 2 weeks, were incorporated.

Five thousand subjects per treatment arm were simulated. Summary statistics were used to define the mean Hgb concentration, along with its associated variabil-

FIGURE 33.5 Schematic of darbepoetin alfa PK/PD model. Darbepoetin alfa concentrations were modeled using a one-compartment model with absorption and lag time. Hgb concentrations were modeled using an indirect response model where darbepoetin alfa concentrations in the central compartment stimulate an increase in RBC production at the precursor stage via an E_{max} model. RBCs then mature at a constant rate (k_t) and manifest as a change in Hgb. Sampling compartments are denoted with lines with solid circles as an arrowhead.

ity, at each sampling point, which were then compared across treatment arms and to actual clinical data collected from 33 patients who were dosed subcutaneously every other week at the dose of 3 μg/kg darbepoetin alfa.

The results are presented in Figure 33.6. No difference was observed in mean change from baseline in Hgb concentrations or in their associated variability over time between any of the groups. The proportions of subjects who were declared positive responders in the observed, simulated weight-based, and simulated fixed-dose treatment groups were 60%, 77%, and 76%, respectively. No statistical difference was observed between these percentages based on their overlapping confidence intervals. Lastly, the percentages of patients requiring transfusion in the weight-based and fixed-dose treatment groups was 21% and 22%, respectively, compared to 16% in the actual observed patients. The higher transfusion rates in the simulation treatment groups were suspected to be due to the objectivity of having a transfusion in clinical practice compared to the yes/no condition defined in the simulation. In summary, these results indicated that dosing per body weight would be an unnecessary complexity and that dosing could proceed based on a fixed dosing regimen, as long as dosing still proceeded based on individual titration to target Hgb values.

Although the results of this analysis have not resulted in a change in the Dosage and Administration section in the Aranesp® product label, based on the results of the simulation, US Oncology, one of the largest oncology consortium in the United States, decided in 2003 to implement a fixed dosing regimen of 200 μg every other week provided Hgb concentrations are maintained at target levels. Thames et al. (26) did a retrospective chart review of US Oncology's practice in 333 patients dosed under these guidelines (174 were previously treated with epoetin alfa and 156 were darbepoetin alfa naive) and found that a "darbepoetin alfa starting dosage of 200 μg every 2 weeks is effective in both naive patients and in those switched from epoetin

FIGURE 33.6 Scatterplot of mean Hgb change from baseline over time in the simulated treatment arms and in 33 patients who were dosed once every other week subcutaneously with 3 μg/mL darbepoetin alfa. Data are reported as the mean ± standard deviation. (Reprinted with permission from Ref. 25.)

alfa." They also found that transfusion rates were about 15% in both the switched and naive treatment groups, which was very near the 16% observed when dosing was done on a weight basis. This simulation illustrates the complexity involved in designing and implementing CTS and illustrates the practical benefit of using CTS in this case—a reduction in the complexity of the dosing regimen.

M&S is not limited to use by the pharmaceutical industry. Frequently, regulatory authorities are using it, sometimes with far-reaching results. Recently, Genta, Inc. (Berkeley Heights, NJ) submitted to the Food and Drug Administration (FDA) a New Drug Application (NDA) for their oligonucleotide Genasense for use in combination with dacarbazine (DTIC) in the first line treatment of patients with advanced melanoma. Genta submitted results from a single, randomized Phase 3 study of DTIC versus DTIC plus Genasense in 771 patients. The primary endpoint was survival, which was not statistically significant ($p = 0.18$). However, the secondary endpoint of progression-free survival (PFS) did show a benefit from 49 days with DTIC to 74 days with DTIC plus Genasense ($p = 0.0003$) using a last-observation carried forward approach to handling censored data. Despite failing to meet their primary endpoint, Genta, Inc. was requesting approval of DTIC in combination with Genasense based on the secondary endpoint of PFS.

In May 2004, the Oncologic Drug Advisory Committee (ODAC) met to discuss Genta's application. Presented at the meeting were the results of a simulation performed by the FDA. The protocol defined disease-free progression as the date on which a scan or measurement was made, not the date of office visit. If patients in one group are assessed at a later date than patents in the other group, the documented date of disease-free progression would be longer in the former group, even if the two groups were equal. Because of a peculiarity in the study design due to drug administration, patients in two treatment groups were not assessed at the same time after the start of the trial. Assessment of patients in the combination arm were slightly delayed compared to patients in the control arm. Using Monte Carlo simulation, the FDA showed that the statistically significant results produced by the sponsor could be an artifact of the trial design and not due to any real drug effect.

This simulation is important for a number of reasons. First, it represents the first real use of simulation by the FDA to question the results from a clinical study. Second, it is informative to listen to a comment made by one of the ODAC members after the FDA presented the results of their simulation. The member stated: "I am sure the 11 or so patients out there still in remission will be disturbed to know that modeling suggests that they shouldn't be there." Clearly, the simulation carried no weight with the physician. This comment highlights a particular problem with simulation in general and that is the credibility gap between modeling and simulation. To most, developing models is one thing, using them is another. The use of simulation in making decisions requires putting one's faith in the model and the assumptions of the simulation. When the 16 members of ODAC were asked whether they believed the observed difference in progression-free survival was real, six members voted "yes." When members were asked whether the difference in response rate and PFS for the combination of DTIC and Genasense versus DTIC alone provided substantial evidence of effectiveness that outweighed its potential for increased toxicity in chemotherapy naive patients with metastatic melanoma, only three members voted "yes." Although the conclusion that members rejected Genta's findings because

of the results of the simulation cannot be definitely made (in fact, their responses would tend to indicate otherwise), it cannot be denied that the committee did in fact reject the sponsor's conclusions and rejected their NDA.

33.9   SUMMARY

CTS sounds daunting. To think that you can simulate a process as complicated as a clinical trial simply sounds crazy. I personally believe that CTS has suffered because of the use of this phrase. But, simulation is nothing more than applied modeling. The principles involved are the same principles that have guided Monte Carlo simulations for the last 50 years. First, a model is needed. Second, the sources of variability in the model parameters must be understood, as does how those parameters are correlated. Third, once the system is defined, an input design must be defined. The process is simulated and the outputs are examined for averages, as well as extrema. There are no black boxes in this process—no smoke and mirrors as I have heard some people call it. The processes involved are the same as those used in other fields, including aerospace, manufacturing, and business. Many software packages aim to simplify the process by making the inner workings more consistent and credible across users, but it is still nevertheless important for the user to understand how these programs work and what to do when they cannot do what is needed.

It remains to be seen what impact CTS will have on drug development—whether it will become an integral part of the process or will become a specialized tool to be utilized on a case-by-case basis. For the former to occur, M&S must gain greater exposure, not among pharmacokineticists and pharmacometricians, but among others impacted by its use like clinicians, project managers, and clinical research associates. While one likes to be recognized by peers, presenting the results of an analysis at meetings geared toward other pharmaceutical scientists, like the American Association of Pharmaceutical Scientists (AAPS), will not necessarily advance the cause of M&S in drug development. Pharmacokineticists must be willing to present their results at more clinically oriented meetings, like the American Society of Clinical Oncology, or more general research oriented programs like the Drug Information Association. Presenting at AAPS is like preaching to the choir; pharmacokineticists are aware of the methodology and want to implement it in their job, but in order to do so must convince the senior leadership in an organization. Only by making them aware of the methodology will M&S be accepted as an integral part of the drug development process.

REFERENCES

1. J. A. DiMasi, R. W. Hansen, and H. G. Grabowski, The price of innovation: new estimates of drug development costs. *J Health Econ* **835**:1–35 (2003).
2. S. Arlington, S. Barnett, S. Hughes, and J. Palo, *Pharma 2010: The Threshold of Innovation*. International Business Consulting Services, London, UK, 2003.
3. P. Van Buskirk, Saving corporations millions: the benefits of modeling. *PCAI*, July/August, 38 (2000).

4. M. D. Hale, Using population pharmacokinetics for planning randomized concentration-controlled trial with a binary response, in *The Population Approach: Measuring and Managing Variability in Response, Concentration, and Dose*, L. Aarons, L. Balant, M. Danhof, M. Gex-Fabry, and U. Gundert-Remy (Eds.). European Commission, Belgium, 1997, pp. 227–235.

5. W. Kang and M. Weiss, Kinetic analysis of saturable myocardial uptake of idarubicin in rat heart: effect of doxorubicin and hypothermia. *Pharm Res* **20**:58–63 (2003).

6. S. M. Eppler, D. L. Combs, T. D. Henry, J. J. Lopez, S. G. Ellis, J.-H. Yi, B. H. Annex, E. R. McCluskey, and T. F. Zioncheck, A target mediated model to describe the pharmacokinetics and hemodynamic effects of recombinant human vascular endothelial growth factor in humans. *Clin Pharmacol Ther* **72**:20–32 (2002).

7. M. Weiss and W. Kang, Inotropic effect of digoxin in humans: mechanistic pharmacokinetic/pharmacodynamic model based on slow receptor binding. *Pharm Res* **21**:231–236 (2004).

8. P. L. Bonate, Clinical trial simulation in drug development. *Pharm Res* **17**:252–256 (2000).

9. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Chapman & Hall, London, 1995.

10. N. H. G. Holford, H. C. Kimko, J. P. R. Monteleone, and C. C. Peck, Simulation of clinical trials. *Annu Rev Pharmacol Toxicol* **40**:209–234 (2000).

11. J. Urquhart, Variable patient compliance as a source of variability in drug response, in *Variability in Drug Response,* G. T. Tucker (Ed.). Elsevier Science, Amsterdam, 1999, pp. 189–198.

12. P. Girard, T. Blaschke, H. Kastrissios, and L. B. Sheiner, A Markov mixed effect regression model for drug compliance. *Stat Med* **17**:2313–2333 (1998).

13. H. Sun, E. O. Fadiran, E. I. Ette, and L. J. Lesko, Regulatory perspectives on clinical trial simulations, in *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis,* P. L. Bonate and D. Howard (Eds.). AAPS Press, Alexandria, VA, 2004.

14. S. M. Ross, *Simulation*. Harcourt/Academic Press, San Diego, CA, 1997.

15. B. D. McCullough, Assessing the reliability of statistical software: Part II. *Am Stat* **53**:149–159 (1999).

16. B. D. McCullough and B. Wilson, On the accuracy of statistical procedures in Excel 97. *Comput Stat Data Anal* **31**:27–37 (1999).

17. B. D. McCullough, Assessing the reliability of statistical software: Part I. *Am Stat* **52**:358–366 (1998).

18. D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 1981.

19. S. K. Park and K. W. Miller, Random number generators: good ones are hard to find. *Commun ACM* **31**:1192–1201 (1988).

20. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*. McGraw-Hill, New York, 2000.

21. G. E. P. Box and M. E. Muller, A note on the generation of random normal variates. *Ann Math Stat* **29**:160–161 (1958).

22. M. Gauthier, Clinical trial simulation. *Appl Clin Trials* **6**:22–25 (1997).

23. M. Hale, W. R. Gillespie, S. Gupta, N. Tuk, and N. Holford, Clinical trial simulation: streamlining your drug development process. *Appl Clin Trials* **5**:35–40 (1996).

24. M. Sale, Modelling and simulation in drug development, promise and reality. *Drug Discovery World* **Spring**:47–50 (2001).

25. N. Jumbe, B. Yao, R. Rovetti, G. Rossi, and A. C. Heatherington, Clinical trial simulation of a 200-microgram fixed dose of darbepoetin alfa in chemotherapy-induced anemia. *Oncology* **16**(Suppl 11):37–44 (2002).

26. W. A. Thames, S. L. Smith, A. C. Scheifele, B. Yao, S. A. Giffin, and J. L. Alley, Evaluation of the US Oncology Network's recommended guidelines for therapeutic substitution with darbepoetin alfa 200 mcg every 2 weeks in both naive patients and patients switched from epoetin alfa. *Pharmacotherapy* **24**:313–323 (2004).

Modeling and Simulation: Planning and Execution

PAUL J. WILLIAMS and JAMES R. LANE

34.1   INTRODUCTION

In recent years modeling and simulation methods have been increasingly used to construct both clinical and preclinical programs and individual studies (1). Simulation has been used to identify optimal study architecture for clinical and preclinical pharmacokinetic (PK), pharmacodynamic (PD), and scale-ups for first-time-in-human (FTIH) studies. Simulation has also been used to create data for communication and graphics, so that the meaning of research can be understood by individuals not involved in pharmacometrics (2). It is conceivable that in the not too distant future as a greater understanding of drug action is realized, late phase clinical development may be minimal or become unnecessary. At that point in time several learning trials will be executed and the results of these trials will be applied to virtual patients to determine the outcomes of drug administration.

Modeling and simulation are especially useful when several critical issues concerning study structure need to be addressed simultaneously. These issues may include dropout rates (especially if related to dose), deviations from protocols by the subject, deviations from protocol by the practitioner, and nested levels of random effects. Such issues have not been addressed when approaches to study structure were based on prior experience, intuition, and empiricism.

Modeling and simulation are a team effort. Several disciplines must be involved in the entire exercise to eventually simulate studies, where there is "buy in" from all stakeholders involved in the drug development process. Prior to a modeling and simulation exercise, appropriate PK, PD, physiology, pathophysiology, and future marketing strategies must be identified; all knowledge must be discovered from all available data (see Chapter 14).

It has been pointed out that modeling and simulation should be guided by clarity, completeness, and parsimony.

Clarity. The report of the simulation should be understandable in terms of scope and conclusions by intended users such as those responsible for committing resources to a clinical trial (see Chapter 37).

Completeness. The assumptions, methods, and critical results should be described in sufficient detail to be reproduced by an independent team.

Parsimony. The complexity of the models and simulation procedures should be no more complex than necessary to meet the objectives of the simulation project. Program codes sufficient to generate models, simulate trials, and perform replication and simulation project level analyses should be retained but there is no need to store simulated trials and analysis results that can be reproduced from these codes (3).

When guided by these principles, simulation in drug development will be streamlined and effective.

The remainder of this chapter deals with the elements of the overall process for an overall project. These include simulation project assessment, project planning, project execution, project reporting, and project utilization.

34.2   EXECUTION OF THE SIMULATION EXERCISE

34.2.1   Simulation Project Evaluation

For simulation to be considered in the drug development scheme, someone involved in the development of a therapeutic agent must realize that it may contribute significantly to the successful development of the agent, by increasing either the probability of approval or the profitability. Simulation can contribute to approval by providing the knowledge necessary for the design of studies that can meet regulatory requirements and studies that increase the likelihood of success by being more powerful. Profitability can be increased by aiding the proper positioning of the agent, shortening timelines to approval, decreasing the number of studies needed for approval, or decreasing individual study cost.

Once the potential value of executing a simulation has been recognized, the feasibility of the project must be determined. To evaluate the feasibility of a project one must assess the current state of knowledge for the agent, evaluate the timeline, and propose a budget. The investigators brochure and all other relevant studies (both clinical and preclinical) must be obtained along with the current proposed study protocol. These must be reviewed and then discussed with those currently involved in the development process. Of note, it is important to involve those who will eventually be marketing the drug to ask what properties of the agent would improve marketability.

Pivotal issues to be addressed when determining feasibility are:

1. Is the project really needed in light of prior knowledge?
2. Will the simulation project add value to the drug?
3. Can the simulation project be executed without compromising current timelines?

4. If the current timeline is compromised, will the simulation add value to the overall drug development process to a such a degree that it is worthwhile?

5. Are the necessary human and monetary resources available for the project?

6. Do existing data and models provide adequate background to make a simulation worthwhile?

34.2.1.1 Project Necessity
The necessity of the project must be evaluated. This can only be done after review of all documents, the study plan, and current data, and discussion with the current development leaders for the agent. It may be that sufficient knowledge has already been created for the project to proceed. It is important to address the trade-offs of doing versus not doing the simulation project. If, for example, a Phase 2 study has already documented efficacy, the dosing strategy, optimal patient selection for Phase 3, and Phase 3 would then be executed as a formality and for assessment of adverse events. In that case a modeling and simulation may not be necessary.

34.2.1.2 Project Completion Without Compromising the Timeline
Very often models have not been developed for use in the ensuing study, and therefore for the simulation project, models must be developed. Developing and validating pharmacometric models can be time consuming. If the simulation was not originally proposed in the development process, the timeline for approval may be compromised. Any delay in development will necessitate that the role of the simulation be scrutinized. However, the delay in the timeline may be defensible, based on an increased certainty of outcomes of the proposed study, the generation of other knowledge that could support registration, or aiding in supporting a go/no-go decision. A strategy should always be in place to execute the needed models as early in the development process as possible. Real-time data collection should be done whenever possible. If blinding is a problem, real-time data collection and assembly should be done offsite. Furthermore, it is possible to develop pharmacometric models for two purposes; one for regulatory approval and the other for use in a simulation exercise. Thus, one may obtain representative data prior to locking of the data set, especially if there is real-time data collection. These data could be used to expeditiously develop and validate a model for the next stage of development.

34.2.1.3 Project Resources
Prior to execution of the project, a serious look at resources required for the simulation project execution must be evaluated. Resources mean more than simply the financial budget. It includes people with the requisite skill set, experience, and knowledge base to execute the project. One must also address whether the necessary computational facilities and softwares are available. If personnel, expertise, or computational facilities are inadequate yet funding is available, then outsourcing the project to a contract research organization (CRO) would be an option. When working with a CRO one must ensure that the contract is very specific regarding the deliverables and timeline. Templates provided to the CRO are of great value when working outside one's own organization. Finally, these contracts need to be executed in a timely manner to ensure projects are completed on time.

34.2.2   Project Plan

34.2.2.1   The End User
Planning begins with the consumer of the simulation outcome in mind. The pharmacometrician has to be conscious of who the end user of the simulation will be and be certain to include the individual who will be making critical decisions, such as a go/no-go decision on the project. At this point the end user of the simulation drives the process; therefore, listening is a crucial quality the pharmacometrician must possess. It is important to play back to the end user what they have said to be certain that there is an understanding of all aspects of the project. The pharmacometrician should not ask questions until the end user has finished making his/her point.

34.2.2.2   Project Purpose
When planning a simulation project, the purpose and intended use of the simulation must be clearly stated. It is the purpose that drives the modeling and simulation process. The clearly stated purpose and intended use of the simulation outcome provide bases for all decision and actions related to the project. The purpose and intended use must be agreed to by all stakeholders in the simulation project.

34.2.2.3   Project Team
The purpose and intended use of the model will determine who the simulation team members are. The team members should cross several disciplines and all pivotal stakeholders need to be identified. The decision maker for future projects and development should be included in the team because it would be extremely frustrating to execute a simulation and, in the end, have it rejected by the end user or by the project decision maker. After the end user and decision makers have been identified, content experts must be added to the team. This would most often include disease experts, statisticians, PM experts (PK and PD experts), computer programmers, and pharmacologists. Individuals from regulatory and marketing are often available for the purpose of identifying what kinds of knowledge would be valuable to gain approval or to optimize the product's position in the marketplace. If there are health-care related cost issues, a member with a background in pharmacoeconomics would be of value. The team should include a simulation team facilitator who should be a simulation expert. The facilitator's role is to oversee the project, conduct expert interviews, gather unbiased assessments, evaluate the models, direct the execution of the project, assign each team member a specific role, and communicate the results. Once the team has been identified, planning of the project can begin.

34.2.2.4   The Project Plan
Clinical trials always have detailed plans and protocols that describe the objectives, hypotheses, assumptions, data collection methods, data analysis methods, and so on. In like manner, the simulation plan describes a simulation process that is agreed to by the simulation team. The plan describes the work to be done, records to be maintained, and reports to be written. This plan should be written in enough detail so that another researcher could pick it up and execute the simulation with corresponding results. The plan must be critiqued and modified if needed. The preparation of the plan also provides an opportunity to evaluate objectives, assumptions, methods, and goals of the project.

The overall objectives of the simulation and how the simulation will be used must be clearly stated in the plan. These will drive model selection and the approach to the execution of the simulation.

If a model must be developed, the model building process must be described, data retrieved, and the model report generated. This process is described in detail in Chapters 8, 14, and 15.

The assumptions of the simulation must be clearly stated. These include the structure of the PK, PD, outcomes, and covariate influence models with the stated values of each parameter. For the execution model, assumptions are made concerning deviations from the protocol, missing data, and patient compliance with the prescribed treatment regimen. The information foundation of the assumptions should be stated; the assumptions are either data and model based, theoretically justified, opinions of domain experts, or conjecture. Premises of greater certainty will often remain unchanged during the simulation process but those of less certainty may be varied to evaluate the final trial for robustness.

The final data analysis approaches for the simulated data must be specified. Software that will be employed should be stated. Standard operating procedures should be referenced.

The design of the simulation study must be stated. For each study the number of replications that will be performed, the factors (e.g., number of subjects, dose enrollment strategies, dropout rates, compliance) and to what degree these factors can be varied, how the robustness of the design will be assessed, and the required informativeness of the study design must be stated. The impact of varying joint factors must also be considered. The number of replications will vary depending on whether only typical outcomes or also atypical outcomes are of interest. Those studies and simulations where atypical or fringe outcomes are of interest will require more replications. An example of a fringe outcome would be the 5th and/or 95th percentiles of a biomarker.

There are three distinct features for each simulation model that must be addressed in the simulation plan. The first are the input–output (IO) models that describe the PK/PD–outcomes models. The inputs here are the rates of drug administration and the outputs are things such as drug concentrations or biomarkers. These IO models should have stochastic elements as part of the model such as between-subject variability and residual variability. It is of primary importance here that the complete probability distribution of the outputs be described in the planning. IO models may be mechanistic or empirical. Mechanistic models attempt to portray the model at the physiological or biochemical level while empirical models simply describe the IO model. Mechanistic models are preferred for simulation as they are more likely to be extrapolated to other studies or drugs in the future.

It is important to incorporate this I/O model parameter uncertainty in the simulation of clinical trials. In order to implement parameter or model uncertainty in the simulation model, the typical values (mean values) of model parameters are usually defined as random variables (usually normally distributed), where the variance of the distribution is defined as standard error squared. The limits of the distribution can be defined at the discretion of the pharmacometrician. For a normal distribution, for example, this would be $\theta \pm 2$ SE, where θ is the parameter. This would include 95% of the simulated distribution. When the simulation is performed, each replicate will have different typical starting values for the system parameters. The

output can then be combined for all replicates and the outcomes of interest studied over a more representative range of variability.

The covariate distribution models, which describe the characteristics of the population (weight, height, sex, race, etc.), must be determined and used for the creation of the study population. The virtual subjects are drawn from a probability distribution that can be one of many types (normal, lognormal, binomial, uniform) but that needs to be described in the study plan. For assignments to sex one must account for what proportion of patients will be female versus male. Furthermore, when creating this population the joint distribution of variables such as height and weight or sex and size must be accounted for. This then leads to the execution model.

The importance of the execution model cannot be overemphasized. The execution model describes how the study is carried out and deviations from the protocol. There are deviations from the protocol that are done by the patient such as refusal to enter the study, dropouts, and patient noncompliance. Other deviations are due to practitioner behavior such as missing data, wrong recording of data, or improper preparation of doses. It must be decided whether the deviations are completely at random or if there is some influencing factor that may result in protocol deviations. For example, would patients experiencing adverse events have a greater tendency to drop out of the study?

As a part of planning, each member of the simulation team must be aware of his/her individual assignments, the deliverables, and timelines. The team members' responsibilities must be stated explicitly and the simulation team facilitator must assure that all members fulfill their responsibilities in a timely manner.

One often overlooked but important part of the plan is for an evaluation of the simulation results when compared to the final study. This must be planned for. Simulation performance criteria must be stated prior to the completion of the final study.

Report templates are often included in the plan. These can be useful as they ensure that important outputs from the simulation project are generated and reported. These templates can include due dates to ensure timely delivery of required inputs and outputs for the simulation project. It is also import to note who will generate and receive the information from these report forms. Often, these templates can be used in future studies either without modification or with only slight modification.

34.2.3 Execution of the Project Plan

A simulation project is an iterative process. Therefore, the team must meet periodically once the simulation is being executed. The simulation is initiated per the simulation protocol. However, some things may change during the execution of the simulation, such as underlying assumptions, or the simulation may show problems that need to be addressed by the team. There may need to be changes in the simulation protocol or the simulation may show flaws in the end study design that should be addressed expeditiously. All affected members of the team should be informed of any proposed changes, especially the end user and decision makers. If there are any changes in the simulation assumptions, then all stakeholders should be notified.

A report of the results of the simulation must be generated. The written report will have the data, assumptions, deviations from the simulation plan with justifications, results, conclusions, recommendations, program scripts, simulation outputs, and supporting literature. The initial draft should be reviewed by the simulation team prior to release to the clinical development team. The final document should be archived.

34.2.4 Applying the Results of the Simulation

The results of the simulation are reported most importantly to the end users and decision makers. The written report is circulated and a team meeting scheduled. The recommendations and reasoning behind any recommendations and options to the recommendations are made concerning the structure of the simulated study. It is important to keep the report and reporting as simple as possible; any presentation with an excessive amount of verbiage or slides is likely to be ignored. The end users and decisions makers will for the most part be interested only in the bottom line of the simulation results. Very often at this point, additional simulations are requested by the clinical development team.

34.3 MISCELLANEOUS POINTS TO CONSIDER

34.3.1 Written Standard Operating Procedures (SOPs)

For organizations that frequently engage in modeling and simulation projects, SOPs are of great value. These should be coupled with templates, practice guidelines, policies, and other similar documents. These documents will help in providing structure to the modeling and simulation process and will expedite the process when repeated. These documents should be reviewed and updated on occasion.

34.3.2 The Problem of Blinded Studies and Data Access

For studies where blinding is a necessity, access to the data is seldom available prior to locking of the database. This causes a problem because the time between locking the database and the writing of the next protocol is not usually sufficient to complete a modeling and simulation project. A solution to this problem is off-site real-time data collection combined with off-site modeling and possibly simulation. An example would be the execution of a Phase 2a study, the results of which will eventually be used to design a Phase 2b study. One may need only a portion of the Phase 2a results (say, only 80% of the Phase 2a data, which would be added to Phase 1 PK/PD data) to develop a PK/PD–outcomes model. Thus, the modeling could begin prior to the data lock. A model would be available at the time of the data lock at the end of Phase 2a. The simulation team could also be selected prior to the data lock. Thus, the simulation could begin in a very timely manner after the data lock and be ready as the protocol is written for the Phase 2b. This type of a workaround must be documented and transparent.

34.4 SUMMARY

A well planned simulation project increases the likelihood of providing meaningful and timely simulation results that will enhance the design and improve the efficiency, robustness, power, and informativeness of preclinical and clinical studies. An increase in the efficiency and power of clinical trials should reduce the number of studies and time needed to complete the drug development process with the resultant reduction in cost of pharmacotherapy to the consumer.

REFERENCES

1. N. Holford, H. Kimko, J. Monteleone, and C. Peck, Simlation of clinical trials. *Annu Rev Pharmacol Toxicol* **40**:209–234 (2000).

2. R. Gieschke and J.-L. Steimer, Pharmacometrics: modeling and simulation tools to improve decision making in clinical drug development. *Eur J Drug Metab Pharmacokinet* **25**(1):49–58 (2000).

3. N. H. G. Holford, M. Hale, H. C. Ko, J.-L. Steimer, L. B. Sheiner, and C. C. Peck, *Simulation in Drug Development: Good Practices*. Available at `cdds.georgetown,edu/research/sddgp723.html` (1999).

Clinical Trial Simulation: Efficacy Trials

MATTHEW M. RIGGS, CHRISTOPHER J. GODFREY, and MARC R. GASTONGUAY

35.1 INTRODUCTION

Expanding on the model-based drug development concept proposed by Sheiner (1), the US FDA in its March 2004 document *Challenge and Opportunity on the Critical Path to New Medical Products* advocates "using simulation software to improve trial design and to predict outcomes." Simulation of a Phase 2b/3 efficacy trial generally occurs during Phase 2 development following proof of concept and dose ranging. However, to say that simulation begins at this point in development is an injustice, or if it is true, a shortcoming of the clinical and scientific development team. Simulation of an efficacy trial should occur as part of a continuum, where accumulated preclinical and clinical information is used to make informative decisions for each next step in development. In this case, simulation will inform the efficacy trial design, conduct, analysis, and interpretation.

In its April 2003 Guidance, *Exposure–Response Relationships—Study Design, Data Analysis, and Regulatory Applications* (`http://www.fda.gov/cder/ guidance/5341fnl.pdf`), the FDA advocates a critical role for exposure–response evaluation in decreasing the uncertainty of drug development. To further emphasize the importance of exposure–response modeling and simulation prior to embarking on large-scale efficacy trials, the FDA Clinical Pharmacology Subcommittee for Pharmacentical Science has suggested early sponsor–FDA meetings (at the end of Phase 2a development) to discuss exposure–response issues (Advisory Meeting Nov. 17/18 2003, `http://www.fda.gov/cder/audiences/acspage/acslist1.htm`).

From this stimulus to learn more about drug effects earlier, a wealth of information should be available for consideration at the point of designing an efficacy trial. Included are preclinical and clinical investigations pertaining to conceptual efficacy and safety endpoints, pharmaceutical and manufacturing considerations for drug product formulation, in vitro and in vivo metabolism characterization, and, in many instances, clinical pharmacology studies exploring potential metabolic interactions and special populations. Efficacy and safety information may include biomarker responses that ideally have been quantified through exposure–response modeling.

The degree of translational relevance of the biomarker(s) to the actual clinical endpoint will be based on prior experience. Information pertaining to adverse event profiles will be emerging from Phase 2 and contextualized through Phase 1 special population and drug–drug interaction (DDI) studies. Additionally, rigorous understanding of "external" information such as disease progression models and comparator profiles will further shape the focus of efficacy trials.

Altogether, the clinical profiles, in conjunction with the nonclinical considerations, should provide a therapeutic range in which to characterize the expected clinical efficacy. Modeling and simulation can offer considerable insight into the design of studies to target this range and ultimately confirm the clinical outcomes. Considerations for performing these simulations are the focus of this chapter. A simulation of a hypothetical efficacy trial for a zidovudine analog in HIV patients is provided as an application and example of many, although not all, of these considerations.

35.2 SIMULATION PLANNING

To begin, all information pertaining to the clinical efficacy trial should be assembled and categorized as possible inputs (known information) or outputs (information that needs to be known) for the trial simulation. This information gathering should involve subject matter experts and key stakeholders of the drug's development (e.g., clinical pharmacology, clinical, statistics, regulatory, operations, and commercial leads). Additional invitees may include other research members (e.g., biology, pharmacology, biopharmaceutics, outcomes research) as determined by the complexity of the input factors (e.g., a thoroughly mechanistic pharmacokinetic/pharmacodynamic (PK/PD) model may benefit from biologist and pharmacologist insight) and output responses (e.g., outcomes research to help define inclusion of clinical utility functions).

As part of this start-up meeting, the identified input factors and output responses should also be assigned a level of precision, either how well it is known (input) or how well it needs to be known (output). Defining this level of uncertainty may be done qualitatively at first, whereas a quantified level of precision will need to be developed for those factors entering into the simulation.

The sections that follow provide examples of both model-based and trial-based input factors and output responses that should be considered for efficacy trial simulations.

35.2.1 Model-Based Input Factors

35.2.1.1 Pharmacokinetic (PK) Model
PK models describe the continuous drug concentration–time course resulting from an administered dose. By doing so for each individual either through (a) intensive collections and standard two-stage PK analyses, or (b) sparse sample collections and population analyses, these continuous descriptions provide a less discrete and so often more informative measure of drug exposure than does dose alone. In concert, mechanistic PK/PD models are being developed more frequently and more congruently during drug development to describe exposure–response relationships (2, 3).

These models often incorporate intermediate biomarker responses. Consequently, trial simulations driven by PK models, rather than more traditional dose–response relationships, will enable more detailed simulations. For example, exposure differences due to interactions, inclusion of special populations, or from dosing regimen or formulation changes may be explored with the PK models driving PD responses. This will place additional emphasis on the modeler to develop reliable PK models using Phase 1 and 2 data that translate into the patient population. Appropriate consideration of covariates, as discussed later, will be an important part of this development.

An additional component of the PK model that may warrant consideration in the simulation is the relative bioavailability of the drug formulation to be used in the efficacy trial. Formulation changes may occur at this point in development, where a suboptimal formulation for commercial-scale manufacture may have been used in previous studies. If such changes have occurred between the dose-ranging study and current design, the relative bioavailability between formulations (and associated 90% CIs) may also be considered for simulation evaluation. A sensitivity analysis (see Section 35.3.1) may be conducted to evaluate whether this effect will be influential on the simulation and/or if additional data may be required to provide acceptable precision.

35.2.1.2 *Exposure–Response (ER or PK/PD) Model*
PK/PD models are becoming less empirical and more mechanistic in nature due to the increasing reliance on biomarkers and the collection of this information earlier and earlier in drug development. The questions being probed by these models through simulation are becoming more focused, including evaluation of exposure differences (4) or even for multiple drugs (2, 5). The former case allows for in silico exploration of formulation (e.g., IR vs. CR) and regimen (e.g., QD vs. BID), or to evaluate effects of exposure differences in special populations or due to DDI. The latter case affords us the opportunity to compare among competing candidates within a development program, or to compare to other available agents.

Implicit to the promising roles of mechanistic PK/PD models is a link to clinical outcomes. A few such examples have been presented, including a model linking blood glucose concentrations and glycosylated hemoglobin (HbA1c) (6), where HbA1c has been linked with progression of nephropathy (7) and retinopathy (8). For atherosclerosis, high serum concentrations of LDL-C have been demonstrated to be a major risk factor for coronary heart disease (CHD) and therapeutic LDL-C lowering (e.g., with HMG CoA reductase inhibitors) has been linked to reduced risk of major coronary events (9).

In addition to exposure–response models for clinical efficacy, the simulation also may include models for safety markers. These may include more immediate or direct effects, such as a drug affecting the QT/QTc interval (10). Although less frequent, longer term effects such as changes in liver function likely are not well defined at this point in development. As known, or potentially expected, such effects may be considered with longitudinal mixture models (11).

Taken together, the ER models for efficacy and safety will define the therapeutic window, where at the point of simulating an efficacy trial, an acceptable separation should exist. However, if multiple markers are being used to determine this separation, there may be a desire to "weigh" some markers more than others. For

example, one marker may be more of a tolerance issue (e.g., moderate incidence of transient nausea upon initiation of therapy) and another a more serious but with a much lower incidence (e.g., edema). Both may warrant inclusion in the simulation but may be assigned different levels of influence through the use of weights. The use of utility functions has been reported to provide this relative balancing (12), which in addition to safety and efficacy parameters, can include any other pertinent input factors as well (e.g., study cost and/or duration).

35.2.1.3   Baseline Distribution Model

Appropriate definition of baseline values for biomarkers or endpoints is critical in setting initial simulation conditions. Both the mean and range of baseline values can have important effects of the projected outcome. This is particularly important for mechanistic models, since the degree of effect is often defined as a function of baseline values. For example, in a typical "indirect" model describing the rate of change in the measured endpoint, R, the assumption may be made that the input rate (k_{in}) is equal to the output rate using a first-order rate constant (k_{out}) and the baseline endpoint measure (R_{base}): $k_{in} = k_{out}R_{base}$ (13). Therefore, baseline models should be evaluated to ensure an appropriate distribution of baseline values is being generated for the patient population of interest. These values also need to be checked against specific inclusion or exclusion criteria for the efficacy study and adjusted as needed.

35.2.1.4   Longitudinal Effect (Disease Progression) Model

The longitudinal course of the targeted disease is of particular interest in many chronic ailments where conditions worsen over time, including Alzheimer's disease, Parkinson's disease, osteoporosis, diabetic nephropathy, and respiratory disease (14). Alternative models have been developed for other disease states where cyclical effects may be observed, such as depression (15) or myelosuppression (5). Included in longitudinal models are descriptions for the natural progression of disease as well as nonpharmacological intervention (e.g., placebo treatment).

   For simulation of an efficacy trial, the mechanism by which the drug affects progression must be considered (e.g., symptomatic or protective). This mechanism may or may not be the same as a comparator agent, which itself may be useful during the simulation. Altogether, the progression model may be used within the simulation to evaluate when clinical endpoints will be measured and the optimal duration of the trial. For example, it may take more or less time (trial duration) than originally anticipated to show separation between treatments depending on the rate of change in the disease progression relative to that affected by standard of care therapy. Detailed discussion of structural considerations for disease progression models is provided in Chapter 21 and elsewhere (14).

35.2.1.5   Covariate Models

Both a drug's PK response and pharmacological response may be influenced by various patient characteristics. Additionally, the progression or extent of disease may be affected by comorbidities. Covariate models attempt to account for and quantify the influence of these factors. For example, a covariate model would be used in simulating PK differences between males and females. Likewise, for a disease progression model of atherosclerosis, or for the overall evaluation of

antihyperlipidemic efficacy, the presence of diabetes raises an individual's cardiovascular risk and results in varying treatment goals for therapy (9). Thus, diabetes may be included as a variable for a lipid efficacy trial simulation where diabetics are part of the study population.

A potential limit at this point in development is that many covariates are still being identified and explored, so the "true" effect is not yet known. Consequently, the precision of the estimated covariate effect may be relatively low. Although it would be advisable to limit the number of covariates included at this stage to those of direct clinical relevance, it is recommended that a "full" model approach (16) be employed. The "full" model would include all covariates of interest with associated mean estimates and precision (e.g., confidence interval (CI) calculated using asymptotic standard errors or bootstrap replication procedures). Collinear covariates should be used with caution as they may affect the precision of the estimates (17).

Using PK differences between males and females as an example, suppose the sex difference was considered "not statistically significant" from Phase 1 and 2 data. However, it may be that the evaluated effect was not "powered" appropriately to rule out a clinically significant difference from the available data. For example, if the resulting 90% CI for the covariate parameter estimate was not well defined (e.g., outside [0.8, 1.25]), then a "no effect" conclusion may not be the most appropriate assumption at this juncture. Therefore, if differences between males and females are of interest to the trial or program outcome, retention of this covariate parameter in the model would be advised. A sensitivity analysis (see Section 35.3.1) assessing the influence of a PK sex effect may be conducted. This would determine whether the point estimate of the effect influences the simulation outcome and how its precision may lend to overall uncertainty in the results.

Covariates are incorporated into the simulation as distributions that are either simulated stochastically or resampled from an existing database (18). Correlation between covariates is handled during stochastic simulations using multivariate distributions with appropriate variance–covariance structure. Alternatively, covariates resampled from a sufficiently large existing database carry all relevant covariates from an individual into a simulated individual and so capture inherent correlation. Regardless of the method, the simulated outputs for covariates need to be checked to ensure that they reflect the expected trial population and are consistent with trial inclusion and exclusion criteria.

35.2.1.6 Compliance Model

Failure to account for nonadherence to study drug administration schedules will lead to biased and imprecise trial simulation outcome measures (19). Models to assimilate compliance often involve a hierarchical Markov model, where the probability for an individual to take a scheduled dose is conditional on whether this individual had taken the previous dose (20, 21). The model may also contain covariates as predictors of compliance. For example, compliance has been shown to be affected by dosing frequency, where an increased frequency (e.g., three times daily vs. once daily) has been associated with worse compliance (22, 23). Alternatively, the consequence of missing a once-a-day dose may have more significant impact on efficacy. PK/PD-based simulations play an important role in understanding the balance of these situations.

In addition to the Markov model, compliance may be modeled using a more simplified model as a mixture (fraction) of patients who are either compliant or noncompliant (all-or-none) (24). Or, similar to drawing covariate distributions from databases of representative populations, a nonmodel-based option for compliance would be to draw from prior compliance data collected from a representative patient population.

35.2.1.7 Study Retention (Dropout) Models

Subjects will drop out of trials for either random (ignorable) reasons or perhaps for a reason attributable to their disease, trial conditions, or other nonignorable factor. Both conditions are important to consider for efficacy trial simulation. In the former case, subjects who drop out (are missing) at random will result in a decrease in total sample size and may affect the study power. In the latter case, nonrandom dropout is considered to be nonignorable in that the reason for dropout is informative to the trial outcome and may bias the results. In the seminal paper by Sheiner (25), an example of nonrandom dropout is presented for an analgesic trial, where those subjects not achieving adequate pain relief were more likely to drop out (i.e., to take rescue medication).

Numerous methods exist for handling dropout, including a recent example by Hu and Sale (26). The reader is referred to a published tutorial (27) for guidance on evaluating the most appropriate method for a particular situation.

35.2.2 Trial-Based Input Factors

Although some of the trial-based input factors will be fixed (e.g., if a design *must be* set a certain way due to unwavering logistics), many of the trial-based factors will be variables for which the simulation will attempt to find an appropriate combination to achieve the trials objectives. These variables are the "what ifs" of the efficacy trial simulation. An attempt to provide a thorough, albeit not all inclusive, list and brief description of trial-based factors to consider for efficacy trial simulation is provided below.

Elemental considerations may be broad comparisons, such as parallel group versus randomized crossover designs. Simulation also may assist in assigning the trial's primary endpoint, where the simulated probabilities of a successful trial for several clinically meaningful outcomes could be used to determine the most appropriate primary endpoint.

Simulation evaluation often considers many numerical factors of a trial design. These include the total number of subjects, the proportion of subjects allocated to the treatment groups, and the number of treatments included (where the range of treatment that is most informative already has been defined by the dose-ranging study). Included in these components may be evaluations to explore effects within specific subpopulations, the inclusion of, and effect in, specific strata within treatment groups, or the impact of other inclusion or exclusion criteria. As discussed earlier, study duration and the number and timing of endpoint measures may also be considered through the trial simulation.

Many of these trial-dependent factors are ultimately evaluated relative to their influence on the statistical power of the study. Additionally, the study data analysis method(s) to be employed may also be a consideration for the simulation.

For example, in the analgesic example cited above (25), a comparison was made between an analysis using the last observation carried forward (LOCF) method and the proposed mixed effects maximum likelihood method. Although this was a retrospective analysis, similar contrasts could be included in the trial's simulation to ascertain the most appropriate analytical methodology to include in the study design (protocol). Other analysis factors for consideration include appropriate correction of variability, where such sources may include differences between sites or regional differences.

35.2.3 Output Responses

The output responses can be categorized as either measurements directly from the trial (e.g., endpoints) or measures of how well the trial is expected to perform (power or probability of success). Either these responses may be evaluated for how well they can be defined for a given study design, or vice versa, they can be defined as needing to be known to a given level of precision, and the study design consequently optimized for this goal. The focal clinical endpoints (need to know) are likely to fall into the latter, while more peripheral endpoints (nice to know) are likely to be assessed as the former.

The simulated output responses provide the measures for which to optimize the design and analysis of the study. In the case of the model-based input factors, their effects are often used to determine how useful (informative) the model(s) and associated parameters are for the output responses and, importantly, which model assumptions may be most critical to the validity of the simulation.

Evaluation of the trial-based input factors is often an iterative process, where various scenarios (e.g., ranges of subject numbers, cohorts, strata, duration) are simulated and the most favorable is carried forward for study planning. The definition of "most favorable" often involves several response factors in addition to the probability of trial success (or failure). Trial costs or other "utility costs" from utility functions may be included in this decision. For example, a larger parallel group study may cost more but finish more quickly than a crossover design. Depending on several other factors, such as within-subject versus between-subject PK and response (efficacy and safety) variances, relative drug effect on disease progression rate, and dropout rates, the designs may have a different probability of success. Therefore, criteria for the cost to benefit ratio for each design would need to be developed and applied to the simulation results. Ideally, such decision trees are formed prospectively, at least in concept, to enable decisive application of the simulation results.

35.2.4 Simulation Team Review of Model and Assumptions

As important as the initial simulation start-up meeting, regular contact with team members must be maintained throughout the trial simulation to ensure continued clarity of the goals, to develop understanding and agreement of the underlying assumptions, and to foster interest, involvement, and ultimately informed application of the simulation results. A second simulation team meeting should ensue once the initial input factors and output responses are identified and initial estimates with accompanying uncertainties are defined. This review should occur prior to

extensive coding, and certainly before beginning the simulation execution. During this review session, all assumptions considered for the simulation should be detailed and agreed upon. A general schematic of the inputs and outputs may be beneficial, with indicators of key inputs, outputs, and assumptions. Once all goals, strengths, benefits, and limitations are agreed upon, the simulations may begin; during which the team is to be provided with regular status updates.

35.3   SIMULATION EXECUTION AND INTERPRETATION

35.3.1   Sensitivity Analysis (SA)

Once the simulation conditions have been properly defined, the simulations are now ready to execute. One important consideration in interpreting simulation results is the sensitivity of these results to underlying assumptions or uncertainty about the simulation model and parameters. Trial simulation outputs should be viewed relative to their sensitivity to model parameter uncertainty. There are two general methods, local and global, for performing SA.

Local SA involves repeated groups of simulations, where in each group a fixed-point perturbation of one parameter is used for the simulations. Trial simulation output metrics are then calculated for each group and the impact on outcome is considered relative to this range of parameter estimates. For example, the degree of sensitivity may be considered by the rate of change in response relative to the unit change in the parameter. This process is then repeated for each parameter of interest. Limitations of the local sensitivity approach are that it only reflects sensitivity to uncertainty in one parameter or assumption at a time. It is therefore inefficient and conclusions about sensitivity are conditional on assumptions of all other parameters.

Global SA is based on simulations where results are conditioned on uncertainty distributions across all parameters. Uncertainty is quantitatively defined for all parameters (models) through the use of appropriate distribution models (28) or using distributions from prior reports or models. The latter method, which does not require an assumed model parameter probability distribution function, may include use of fuzzy set theory (29) or the use of bootstrapped estimates from previous estimations. Monte Carlo methods are required to simulate from the uncertainty distributions at the intertrial level. This usually requires one set of simulations with a large number of replicates. The number of trial replicates is discussed in Chapter 33, where this number may need to be further increased for the global SA.

One benefit of global versus local SA is that by running only one set of simulations, the sensitivity of simulation outcome(s) to simulation parameter uncertainty (assumptions) can be viewed over a continuous range of parameter uncertainty because the sensitivity to uncertainty is incorporated in all model parameters simultaneously. Examples of simulation with uncertainty are available for physiologically based PK models (29, 30) and clinical trial simulations (31, 32).

In practice, a blend of local and global SA approaches may be employed. Quantitative uncertainty in the form of standard errors may be available only for some model inputs. Multiple executions of the simulation to characterize the global SA of the quantified parameters may be undertaken conditional on a series of fixed-

point perturbations of another parameter. This "semiglobal" SA would determine whether the findings about global sensitivity change given the uncertainty in the locally perturbed parameter.

35.3.2 Simulation Team Review of Sensitivity Analysis and Impact on Assumptions

The SA will identify which parameters are most influential on the simulation outcome and consequently which assumptions are either reasonable or in doubt. This provides another decision point for the overall trial simulation, and so a meeting with the simulation team should be convened. This meeting should cover the impact of the SA on the trial (and/or program) planning and design. As an example discussion point, if the SA reveals previously unexpected, and possibly undesirable, response ranges in a certain patient subpopulation, then the team would discuss whether (a) to proceed with the current design, given the risk of unexpected or undesirable outcomes, which has now been quantified; (b) the model assumptions for that subpopulation are appropriate; (c) more data be collected to reduce the uncertainty surrounding this patient population before executing the trial; (d) the current trial design requires modification to make it more robust to potential differences in this subpopulation; or (e) this subpopulation be excluded altogether if they are not of interest in the current trial.

The SA may also provide rationale for simplifying a model. This may occur if the outcome is shown to be robustly tolerant to wide ranges of parameter uncertainty, which may allow for the removal of some parameters and thus lead to a more parsimonious simulation model. Conversely, the SA may reveal that insufficient information currently exists to define a precise or reliable range of trial outcomes. In this latter case, either more time may be required to obtain additional informative experimental data and thus reduce the uncertainty to an acceptable range, or separate sets of plausible assumptions may need to be considered and subsequently tested for their own sensitivity. Such decisions need buy-in from the subject matter experts and should be considered in the full context of the development program.

35.3.3 Simulation Recommendations

Following discussion and acceptance of the SA results, including both model-based and trial-based input factor adjustments, the efficacy trial simulations may proceed as planned. For each possible trial design, the appropriate input factors and output responses are simulated and results are compared to determine the most appropriate design. As discussed previously, this final decision likely will not only be based on a specific *p*-value or trial power, but will also include valuations based on trial duration, monetary cost, or information gained or lost toward continuing development goals (e.g., an overall measure of clinical utility).

35.4 EFFICACY TRIAL SIMULATION EXAMPLE

A simulation of a hypothetical efficacy trial for a zidovudine analog (ZDVA) in HIV patients was completed to evaluate the probability of a successful Phase 3 trial if Phase 2b was skipped, given the Phase 2a results and prior knowledge from

a marketed competitor. The goal of this simulation was to evaluate the probability of trial success under a predefined trial design while examining the sensitivity of this probability estimate to underlying assumptions. Included in this example is an assessment of the effect of uncertainty (from none through varying degrees) on the trial outcome.

35.4.1 Model-Based Inputs

Input factors, as defined in Table 35.1 and presented schematically in Figure 35.1, included an adherence model, PK and PK/PD models, placebo response and disease progression models, and both random and nonignorable dropout. The PK/PD model was used to simulate patient survival times (or censored events) for each individual as a function of drug exposure, which was defined as the average steady-state drug plasma concentration (C_{avg}). C_{avg} was modeled as a function of clearance (CL), which depended on hepatic disease (HEP), weight (WT), and methadone use (METH). Dosing compliance/adherence was modeled with a bimodal distribution (high and low compliance groups). A hazard function for dropout due to inefficacy included terms for placebo dropout, effect of CD4+ count, and effect of ZDV. Random dropout was also simulated. The full model is provided in the chapter appendix.

"Uncertainty" distributions were defined for all parameters including typical PK, PD parameters, covariate effects, and interindividual and residual variance parameters. These distributions were derived from the variance–covariance matrix of the estimates obtained from a prior analysis, and from a review of prior knowledge and published results.

Separate simulations were performed with no, low, moderate, or high degrees of uncertainty to explore the effect of uncertainty on the probability of trial success using a global SA. Uncertainty, expressed as %CV for each parameter, was set for the fixed model effects at 0% (none), 10% (low), 35% (moderate), or 50% (high).

TABLE 35.1   Description of Simulation Model Parameters Used for the Zidovudine Analog Efficacy Trial Simulation

Name	Description	Name	Description
CD4	Mean CD4+ count	$METH_{CL}$	Methadone effect on CL
CD4PT	CD4+ breakpoint for beneficial effect	PrFEM	Probability of female
CD4SLP	Effect of CD4 count on hazard	PrHEP	Probability of hepatic disease
CLBASE	Typical clearance	PrHICOMP	Probability of high compliance
HAZP	Placebo risk for dropout (hazard)	PrLOWCOM	Probability of low compliance
HAZR	Random dropout hazard	PrMETH	Probability of methadone use
HEP_{CL}	Effect of hepatic disease on CL	SEX_{CL}	Sex effect on CL
HICOMP	Mean high compliance	WT_{FEM}	Mean weight for females
IC50	Drug potency	WT_{MALE}	Mean weight for males
LOWCOMP	Mean low compliance	ZDVSL	Maximum drug effect on hazard

FIGURE 35.1   Schematic of zidovudine analog efficacy trial simulation components.

Uncertainty for random effects was defined as high for all evaluations. The high level of uncertainty for random effects represented a "worst case" scenario, but also was representative of the greater uncertainty in these parameters at this stage in development relative to the fixed effects parameters. In an actual trial simulation, prior estimates of these uncertainties could be used, and possibly inflated to accommodate the additional parameter uncertainty associated with extrapolation to a new trial scenario or population. To illustrate a local SA, the parameter describing the maximum drug effect on the hazard parameter (ZDVSL) was fixed at values ranging from 0.25 to 1.0.

35.4.2   Trial-Based Inputs

Two thousand patients were to be randomly assigned to placebo, ZDVA 500 mg or ZDVA 1500 mg, daily. Follow-up visits were planned every 28 days for 2 years. The number of "survivors" was defined as the number of patients remaining in the study at each observation time.

35.4.3   Simulation Execution and Analysis

Simulations with and without varying degrees of uncertainty were performed, as described above. Simulation with uncertainty can be implemented in a variety of programs with Monte Carlo simulation capabilities. In this example, simulations were carried out using S-Plus® (Insightful, Seattle, WA) and NONMEM® (Globo-Max LLC, Ellicott City, MD). PROC PHREG in SAS® (SAS Institute Inc., Cary, NC) was used for survival analyses. Local regression plots were created with S-Plus and the LOCFIT library.

35.4.4   Simulation Results and Recommendation

The effect of the degree of uncertainty on trial success (power) is provided in Table 35.2. A total of 500 replicate simulations were performed for each extreme of uncertainty (none and high). Fewer ($n = 100$) replicates were performed for the intermediate levels (low and medium), based both on practical computational limitations and an observation of relative stability in the results after 100 replications. More ($n = 5000$) replicates were performed for graphical clarity in the local regression plots. The results show that failure to account for uncertainty in the model would result in an overestimate of the trial power and thus a falsely optimistic design.

TABLE 35.2 Effect of Uncertainty Level on the Estimate of Trial Power Using a Global Sensitivity Analysis for the Zidovudine Analog Efficacy Trial Simulation

Uncertainty	Successful Trials[a] (%)
None	94.4% ($n = 500$)
Low (10%)[b]	93% ($n = 100$)
Medium (35%)[b]	86% ($n = 100$)
High (50%)[b]	82.8% ($n = 500$)

[a] Results reflect n simulated trials of 2000 patients.
[b] Uncertainty in fixed effect parameters is approximate %CV. Uncertainty in random effect parameters was high for all simulations.

FIGURE 35.2 Results of global SAs from the zidovudine analog efficacy trial simulation. Parameters displayed are for drug clearance (CLBASE), drug potency (IC50), placebo risk on the dropout hazard (HAZP), and maximum drug effect (ZDVSL) on the dropout hazard. The trial power (probability of success) is plotted relative to the parameter value for each of the 5000 replicate simulations. The solid lines indicate the local logistic regression (locfit) and the dotted lines provide the 95% confidence interval for the locfit.

A global SA for the simulated outcome, probability of a successful differentiation from placebo, was not sensitive to uncertainties in CL and IC50 parameters (Figure 35.2).

Simulation conclusions were sensitive to assumptions about ZDVSL and HAZP parameter values (Figures 35.2), and a wide range of possible outcomes was evident.

In other words, at the hypothetical doses explored (500 and 1500 mg), uncertainties about exposure (CL) and potency (IC50) were not important contributors to the uncertainty of response, whereas uncertainties about the maximal effect (ZDVSL) and placebo dropout did contribute notably to the trial power.

An additional benefit of the global SA is that it provided the interdependence of the ZDVSL and HAZP parameters on the trial outcome (Figure 35.3), illustrating a benefit of global versus local SA.

The results of a local SA using the ZDVSL parameter are provided in Table 35.3. These results are similar to the global SA, but are conditioned on the fixed estimates of the other parameters.

As shown with both global and local SA, the probability of a successful trial was less than 80% across a large range of uncertainty in the parameters (Figures 35.2 and 35.3 and Table 35.3). Given prespecified criteria of ≥80% power, the

FIGURE 35.3   Results of the global SA from the zidovudine analog efficacy trial simulation displaying the multidimensional effect of parameter uncertainty on trial power (probability of success). Parameters displayed are placebo risk on the dropout hazard (HAZP) and maximum drug effect (ZDVSL) on the dropout hazard.

**TABLE 35.3   Effect of Single Parameter (ZDVSL[a])
Uncertainty on the Estimate of Trial Power Using a Local
SA for the Zidovudine Analog Efficacy Trial Simulation**

Fixed Value of ZDVSL	Successful Trials[b] (%)
0.25	30.6%
0.5	70.4%
0.735	93.0%
1.0	99.0%

[a] ZDVSL represents the maximum drug effect on the dropout hazard.
[b] Results reflect 500 simulated trials of 2000 patients.

simulation results represented an unacceptable risk of proceeding with the current design, given the current state of knowledge. More robust trial designs might have proved to be impractical. Therefore, before proceeding with the current design, it would be recommended to refine estimates of these influential model components. The overall recommendation from this hypothetical example would be to run the Phase 2b trial, rather than skipping directly into Phase 3.

35.5   SUMMARY

Returning to the FDA March 2004 document *Challenge and Opportunity on the Critical Path to New Medical Products*, "the current medical product development path is becoming increasingly challenging, inefficient, and costly. A new product development toolkit—containing powerful new scientific and technical methods such as . . . computer-based predictive models . . . is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product." Efficacy trial simulation satisfies this initiative by offering a tangible benefit to drug development through an a priori in silico evaluation of trial design performance. Consistent development and application of this approach to program planning will lead to more informed study designs and more focused outcomes, while minimizing costs and time. To appropriately apply this approach, development and management teams must champion its benefit and provide dedicated resources, as well as commit project team time to correctly interweave simulation into the development process.

ACKNOWLEDGMENT

REFERENCES

1. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

2. M. O. Karlsson, T. Anehall, L. E. Friberg, A. Henningsson, C. Kloft, M. Sandstrom, and R. Xie, Pharmacokinetic/pharmacodynamic modelling in oncological drug development. *Basic Clin Pharmacol Toxicol* **96**:206–211 (2005).

3. D. R. Huntjens, M. Danhof, and O. E. la Pasqua, Pharmacokinetic-pharmacodynamic correlations and biomarkers in the development of COX-2 inhibitors. *Rheumatology* **44**:846–859 (2005).

4. S. Chabaud, P. Girard, P. Nony, and J. P. Boissel, Clinical trial simulation using therapeutic effect modeling: application to ivabradine efficacy in patients with angina pectoris. *J Pharmacokinet Pharmacodyn* **29**:339–363 (2002).

5. L. E. Friberg, A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson, Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol* **20**:4713–4721 (2002).

6. T. M. Post, W. de Winter, J. DeJongh, I. Moules, R. Urquhart, D. Eckland, and M. Danhof, Treatment efficacy of combination-therapy based on a mechanistic characterisation of disease processes in type 2 diabetes mellitus over a two-year period. *Abstracts of the Annual Meeting of the Population Approach Group in Europe. PAGE* **14**:Abstr 757 (2005).

7. P. Hovind, P. Rossing, L. Tarnow, U. M. Smidt, and H. H. Parving, Progression of diabetic nephropathy. *Kidney Int* **59**:702–709 (2001).

8. O. Brinchmann-Hansen, K. Dahl-Jorgensen, L. Sandvik, and K. F. Hanssen, Blood glucose concentrations and progression of diabetic retinopathy: the seven year results of the Oslo study. *BMJ* **304**:19–22 (1992).

9. S. M. Grundy, J. I. Cleeman, C. N. Merz, H. B. Brewer, Jr., L. T. Clark, D. B. Hunninghake, R. C. Pasternak, S. C. Smith, Jr., and N. J. Stone, Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Arterioscler Thromb Vasc Biol* **24**:e149–e161 (2004).

10. D. M. Jonker, L. A. Kenna, D. Leishman, R. Wallis, P. A. Milligan, and E. N. Jonsson, A pharmacokinetic-pharmacodynamic model for the quantitative prediction of dofetilide clinical QT prolongation from human ether-a-go-go-related gene current inhibition data. *Clin Pharmacol Ther* **77**:572–582 (2005).

11. K. G. Kowalski, L. McFadyen, M. M. Hutmacher, B. Frame, and R. Miller, A two-part mixture model for longitudinal adverse event severity data. *J Pharmacokinet Pharmacodyn* **30**:315–336 (2003).

12. G. Graham, S. Gupta, and L. Aarons, Determination of an optimal dosage regimen using a Bayesian decision analysis of efficacy and adverse effect data. *J Pharmacokinet Pharmacodyn* **29**:67–88 (2002).

13. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**:457–478 (1993).

14. P. L. Chan and N. H. Holford, Drug treatment effects on disease progression. *Annu Rev Pharmacol Toxicol* **41**:625–659 (2001).

15. N. Holford, J. Li, L. Benincosa, and M. Birath, Population disease progress models for the time course of HAMD score in depressed patients receiving placebo in antidepressant clinical trials. *Abstracts of the Annual Meeting of the Population Approach Group in Europe. PAGE* **11**:Abstr 311 (2002).

16. F. E. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York, 2001, pp. 79–83.

17. J. Ribbing and E. N. Jonsson, Power, selection bias and predictive performance of the population pharmacokinetic covariate model. *J Pharmacokinet Pharmacodyn* **31**:109–134 (2004).

18. D. R. Mould, Defining covariate distribution models for clinical trial simulation, in *Simulation for Defining Clinical Trials*, H. Kimko and S. Duffull (Eds.). Marcel Dekker, New York, 2001, pp. 31–54.

19. R. R. Bies, M. R. Gastonguay, K. C. Coley, P. D. Kroboth, and B. G. Pollock, Evaluating the consistency of pharmacotherapy exposure by use of state-of-the-art techniques. *Am J Geriatr Psychiatry* **10**:696–705 (2002).

20. P. Girard, T. F. Blaschke, H. Kastrissios, and L. B. Sheiner, A Markov mixed effect regression model for drug compliance. *Stat Med* **17**:2313–2333 (1998).

21. P. Girard, L. B. Sheiner, H. Kastrissios, and T. F. Blaschke, Do we need full compliance data for population pharmacokinetic analysis? *J Pharmacokinet Biopharm* **24**:265–282 (1996).

22. J. A. Cramer, M. M. Amonkar, A. Hebborn, and R. Altman, Compliance and persistence with bisphosphonate dosing regimens among women with postmenopausal osteoporosis. *Curr Med Res Opin* **21**:1453–1460 (2005).

23. J. J. McNabb, D. P. Nicolau, J. A. Stoner, and J. Ross, Patterns of adherence to antiretroviral medications: the value of electronic monitoring. *AIDS* **17**:1763–1767 (2003).

24. A. J. O'Malley and S. L. Normand, Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. *Biometrics* **61**:325–334 (2005).

25. L. B. Sheiner, A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther* **56**:309–322 (1994).

26. C. Hu and M. E. Sale, A joint model for nonlinear longitudinal data with informative dropout. *J Pharmacokinet Pharmacodyn* **30**:83–103 (2003).

27. J. W. Hogan, J. Roy, and C. Korkontzelou, Handling drop-out in longitudinal studies. *Stat Med* **23**:1455–1497 (2004).

28. A. M. Law and W. D. Kelton, Selecting input probability distributions, in *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York, 2000, pp. 292–401.

29. I. Nestorov, I. Gueorguieva, H. M. Jones, B. Houston, and M. Rowland, Incorporating measures of variability and uncertainty into the prediction of in vivo hepatic clearance from in vitro data. *Drug Metab Dispos* **30**:276–282 (2002).

30. D. Farrar, B. Allen, K. Crump, and A. Shipp, Evaluation of uncertainty in input parameters to pharmacokinetic models and the resulting uncertainty in output. *Toxicol Lett* **49**:371–385 (1989).

31. W. R. Gillespie, M. Sale, and M. R. Gastonguay, Modeling uncertainty in clinical trial simulation: a Bayesian approach. *Clin Pharmacol Ther* **65**:Abstr, PIII-21 (1999).

32. H. Kraiczi and M. Frisen, Effect of uncertainty about population parameters on pharmacodynamics-based prediction of clinical trial power. *Contemp Clin Trials* **26**:118–130 (2005).

APPENDIX 35.1 NONMEM CODE FOR EFFICACY TRIAL SIMULATION

The code does not include parameters for uncertainty. Parameter ZDVSL noted in chapter text is termed BZDV in the NONMEM code.

```
$PRED
  ;FIRST RANDOM SOURCE IS FOR ETAS
  IF (ICALL.LE.1) RETURN ;IF BEGINNING OF RUN OR PROBLEM
    Y=0
  ;2ND RANDOM SOURCE (UNIFORM)
  IF (ICALL.EQ.4) THEN ; FOR SIMULATION OF SURVIVAL TIMES GIVEN
HAZARD FUNCTION
  CALL RANDOM (2,R) ;RANDOM SEED FROM UNIFORM (0,1)
  X=R
ENDIF
;2ND RANDOM SOURCE (UNIFORM)
IF (ICALL.EQ.4) THEN ; FOR SIMULATION OF SURVIVAL TIMES GIVEN
HAZARD FUNCTION
  CALL RANDOM (2,R) ;RANDOM SEED FROM UNIFORM (0,1)
  X2=R
ENDIF
```

```
;NOW TAKE CARE OF PATIENT DEMOGRAPHICS
;FOR PROBABILITIES, USE TRANSFORMATION PR=EXP(THETA)/(1+EXP(THETA))
TO CONSTRAIN 0-1
  IF (ICALL.EQ.4.AND.NEWIND.NE.2) THEN ;FOR $SIM AND 1ST RECORD OF
EACH IND
  ;EACH RECORD IN THIS DATA SET IS THE
  ;FIRST RECORD FOR THAT INDIVIDUAL -
  ;ONLY 1 RECORD PER INDIVIDUAL
  ;2ND RANDOM SOURCE (UNIFORM) - FOR HEPATIC DISEASE (3% OF
PATIENTS)
  CALL RANDOM (2,R)
  RH=R
  HEP=0
  PRHD=EXP(THETA(10))/(1+EXP(THETA(10)))  ;PROBABILITY OF HEPATIC
DISEASE - THIS WILL HAVE UNCERTAINTY
  IF (RH.GT.(1-PRHD)) HEP=1 ;HEPATIC DISEASE


  ;2ND RANDOM SOURCE (UNIFORM) - FOR SEX (79% MALE, 21% FEMALE)
  CALL RANDOM (2,R)
  RS=R
  SEX=0 ;MALES
  PRFE=EXP(THETA(11))/(1+EXP(THETA(11)))
    IF (RS.GT.(1-PRFE)) SEX=1 ;FEMALES


  ;2ND RANDOM SOURCE (UNIFORM) - FOR METHADONE USE (4.4% OF
PATIENTS)
  CALL RANDOM (2,R)
  RM=R
  METH=0
  PRME=EXP(THETA(12))/(1+EXP(THETA(12))) ;PROBABILITY OF METHADONE
USE IS 0.044
    IF (RM.GT.(1-PRME)) METH=1 ;ON METHADONE


  ;3RD RANDOM SOURCE (NORMAL) - FOR WEIGHT (NORMAL WITH MEAN OF
73 FOR MALES,
  ; 60 FOR FEMALES CV=20%, RANGE=40-100)
  WT=1
  DO WHILE (WT.LT.40.OR.WT.GT.100)
    CALL SIMETA (ETA)
  LBWT=40
  UBWT=100
  RGWT=UBWT-LBWT
  WT1=LBWT+RGWT*EXP(THETA(13))/(1+EXP(THETA(13)))
  WT2=LBWT+RGWT*EXP(THETA(14))/(1+EXP(THETA(14)))
  TVWT=(WT1*(1-SEX) + WT2*SEX)
  WT=TVWT*(1 + ETA(7))
ENDDO
```

```
  ;1ST RANDOM SOURCE (NORMAL) - FOR CD4+ COUNT (NORMAL WITH MEAN
OF 348, SD=90, MAX=500)
  CD4=0
  CD=501
  DO WHILE (CD.LE.0.OR.CD.GT.500)
    CALL SIMETA (ETA)
    LBCD=0
    UBCD=500
    RGCD=UBCD-LBCD
    TVCD=LBCD+RGCD*EXP(THETA(15))/(1+EXP(THETA(15)))
    CD=TVCD+ETA(6)
    CDPT=EXP(THETA(16)) ; BREAKPOINT FOR SIGNIFICANT CD4 EFFECT ON
EFFICACY
    IF (CD.GE.CDPT) CD4=1
    ENDDO

  ;2ND RANDOM SOURCE (UNIFORM) - FOR DOSE ASSIGNMENT
  CALL RANDOM (2,R)
  RD=R
  DOSE=0
  IF (RD.GE.0.33.AND.RD.LT.0.67) DOSE=500*28
  IF (RD.GE.0.67) DOSE=1500*28 ; DOSE FOR 4 WEEKS

  ;USE 2ND RANDOM SOURCE FOR COMPLIANCE MODEL - COMPLIANCE IS
RANDOM
  ;COMPLIANCE IS MODELED AS A BINOMIAL DISTRIBUTION WITH 90% HI
COMPLIANCE MEAN=80%, SD=10
  ;AND 10% LOW COMPLIANCE COMPL=40%, SD=5 (MIN=0, MAX=100%)
  ;CALL 2ND RANDOM SOURCE (UNIFORM) - FOR TYPE OF COMP (90% HIGH,
10% LOW)
  CALL RANDOM (2,R)
  RC=R
  CTYP=0 ;HIGH COMPLIANCE
  PRHI=EXP(THETA(17))/(1+EXP(THETA(17)))
    IF (RC.GT.PRHI) CTYP=1 ;LOW COMPLIANCE
  COMP=1.1
  DO WHILE (COMP.LT.0.OR.COMP.GT.1)
    CALL SIMETA (ETA)
    CMP1=EXP(THETA(18))/(1+EXP(THETA(18)))
    CMP2=EXP(THETA(19))/(1+EXP(THETA(19)))
    COMP=(CMP1+ETA(8))*(1-CTYP) + (CMP2+ETA(9))*CTYP
  ENDDO
  ADOS=DOSE*COMP ;ACTUAL DOSE=ASSIGNED DOSE*COMPLIANCE

  ENDIF

; CL=1.3+/-0.3 L/KG/HR CL IN L/4 WEEKS = CL L/KG/HR*24HRS/DAY
*28DAYS*WT
```

```
; USE EXP MODEL AND 23% INTERINDIVIDUAL CV INSTEAD
  TVCL=EXP(THETA(1))
  CL1 = TVCL *EXP(ETA(1))
  CL2=(EXP(THETA(2)))**HEP
  CL3=(EXP(THETA(3)))**SEX
  CL4=(EXP(THETA(4)))**METH
  CL = CL1*WT*24*28*CL2*CL3*CL4
  CAVG = ADOS/CL ;(AVERAGE CSS)
  TVIC=EXP(THETA(5))
  IC50 = TVIC*EXP(ETA(2))
  TVHP=EXP(THETA(6))
  HP = TVHP*EXP(ETA(3))
  TVBC=EXP(THETA(7))
  BCD4 = TVBC*EXP(ETA(4))
  TVZD=EXP(THETA(8))
  BZDV = TVZD*EXP(ETA(5))


;HAZARD MODEL FOR BASELINE CD4 EFFECT ON EFFICACY
  HAZ = HP*EXP(-BCD4*CD4 + (-BZDV*CAVG/(CAVG+IC50)))


;IN THIS CASE, THE HAZARD FUNCTION IS CONSTANT OVER TIME FOR EACH
INDIVIDUAL
;THE OBSERVATION IS THE EVENT TIME
;X IS RANDOM NUMBER DRAWN FROM A UNIFORM (0,1) DISTRIBUTION
  CTIME=24 ;24 MONTHS
  OBS1=LOG(1/(1-X))/(HAZ)
  CEN1=1 ;FLAG INDICATING THAT THE EVENT DID HAPPEN BEFORE
CENSORING TIME
  IF (OBS1.GT.CTIME) THEN
    CEN1=0 ;EVENT HAPPENS AFTER CENSORING TIME
    OBS1=CTIME
  ENDIF
  ;RANDOM DROPOUT EVENT: Pr(T<=t) BY RANDOM CHANCE = 0.167%
  HR = EXP(THETA(9))/(1+EXP(THETA(9)))
  OBS2=LOG(1/(1-X2))/HR
  CEN2=1
  IF (OBS2.GT.CTIME) THEN
    OBS2=CTIME
    CEN2=0
  ENDIF
  OBS=OBS1
  IF (OBS2.LT.OBS1) OBS=OBS2
  CENS=0
  IF (CEN1+CEN2.GT.0) CENS = 1
  RNDP=0
  IF (OBS2.LT.OBS1) RNDP=1
TRL=IREP ;TRIAL NUMBER
```

```
$THETA ;INITIAL ESTIMATES FOR THETAS
  ;SOME PARAMETERS WERE LOG OR LOGIT TRANSFORMED TO CONSTRAIN
PARAMETERS
  ;THIS IS NECESSARY BECAUSE DRAWS FROM PRIOR DISTRIBUTION ARE NOT
TRUNCATED AT REALISTIC VALUES
  (0.26)   ;1. LN(CL =1.3 L/KG/HR )
  (-1.39)  ;2. LN(FRACTIONAL EFFECT OF HEP ON CL=0.25)
  (-0.22)  ;3. LN(FRACTIONAL EFFECT OF SEX ON CL=0.8)
  (-0.51)  ;4. LN(FRACTIONAL EFFECT OF METH ON CL=0.6)
  (-4.34)  ;5. LN(ZDV IC50 FROM LITERATURE=0.013)
  (-4.83)  ;6. LN(HP - HAZARD FOR PLACEBO AND CD4<300 OVER 28
DAYS=0.008)
  (-0.53) ;7. LN(BCD4 - BASELINE CD4 EFFECT ON HAZARD WHEN CD4>=300
= 0.59)
  (-0.693) ;8. LN(BZDV - ZDV EFFECT ON HAZARD=0.735)
  (-6.39)  ;9. LN(HR HAZARD FOR RANDOM DROPOUT, ASSUMING 0.167%
CHANCE OF DROPOUT/28 DAYS)
  (-3.476) ;10. PROBABILITY OF HEPATIC DISEASE=0.03
  (-1.325) ;11. PROBABILITY OF PATIENT BEING FEMALE=0.21
  (-3.076) ;12. PROBABILITY OF METHADONE USE=0.044
  (0.2) ;13 (73) MEAN MALE WT
  (-0.69) ;14 (60) MEAN FEMALE WT
  (0.83) ;15 (348) MEAN CD4 COUNT
  (5.7) ;16 LN(300) BREAKPOINT FOR CD4 EFFECT
  (2.20) ;17 PROBABILITY OF HIGH COMPLIANCE=0.9
  (1.39) ;18 (0.8) 80% COMPLIANCE=HIGH
  (-0.41) ;19 (0.4) 40% COMPLIANCE=LOW

$OMEGA ;INITIAL ESTIMATES FOR OMEGA
  0.053 ;1. CL
  0.09 ;2. IC50
  0.09 ;3. HP
  0.09 ;4. BCD4
  0.09 ;5. BZDV
  8100 ;6. CD4
  0.04 ;7. WT
  100 ;8. HI COMPL
  25 ;9. LOW COMPL

$SIMULATION (78925 NORMAL NEW) (8795 UNIFORM)(15445 NORMAL)
  ONLYSIM SUBPROBLEMS=500
```

PHARMACOMETRIC SERVICE AND COMMUNICATION

Engineering a Pharmacometrics Enterprise

THADDEUS H. GRASELA and CHARLES W. DEMENT

36.1   INTRODUCTION

36.1.1   Transition to Model-Based Development

Pharmacometrics is no longer a hobby for scientists who want to learn the arcane language of mixed effects modeling. The evidence for this transition from pastime to critical path can be found in the change in emphasis from straightforward population pharmacokinetic modeling to complex population pharmacokinetic and pharmacodynamic modeling and simulation activities that yield multifaceted exposure–response-based characterizations of safety and efficacy. This transition is further evidenced in the myriad ways that exposure–response analyses are being used to select and justify doses for Phase 3 programs, support product labeling and differentiation, and influence drug development and regulatory decision making.

This transition is also marked by the challenges faced by pharmacometricians conscientiously working within the existing paradigm to support the current empirically driven development process. War stories abound that describe the difficulties pharmacometricians face in delivering timely and actionable results from their analyses. The data required for pharmacometric analysis are often not available until the primary safety and efficacy analyses have been completed. Data assembly and scrubbing are remarkably time consuming and can result in high discard rates and delays in completing modeling and simulations. One can still encounter significant resistance to the use of modeling and simulation results in development program decision making, and opportunities for collaboration, creative thinking, and synthesis of knowledge may be sacrificed because of urgent timelines (1).

These challenges are not just annoyances that committed pharmacometricians must be expected to overcome. They are symptoms of a deeper problem—an immature process capability for performing modeling and simulation—resulting from the current, ad hoc implementation wherein modeling and simulation activities are

piggybacked onto traditional development programs. These symptoms notwithstanding, the growing reliance on modeling and simulation in decision making is forcing important and urgent changes to occur in the nature of the work of a pharmacometrics service. Moreover, the role of pharmacometrics in the larger drug development enterprise is changing very rapidly because of the need to make the parent drug development process more efficient and effective. This larger need is driving a shift from a totally empirical paradigm to a more formal, model-driven one. Accordingly, what it takes to establish a pharmacometrics service (or sustain and strengthen an existing one) is also changing significantly from what it has required in the past.

Effectively meeting the challenges one confronts in establishing a service in the face of this shift in the developmental center of gravity mandates the use of systematic and rigorous methods to consciously and deliberately "engineer" the service, or "reengineer" it, as the case may be. That is, successfully deploying a complex technical service such as pharmacometric analysis in today's complex and fluid environment mandates prior development of robust enterprise designs and effective implementation programs. Doing this right, in turn, requires use of a process that is designed explicitly to do this sort of job—the systems engineering process (2).

Systems engineering can be defined as an interdisciplinary approach encompassing the entire effort to evolve and verify an integrated set of system, people, product, and process solutions that satisfy stakeholder needs. In the context of a pharmacometrics service, systems engineering encompasses (a) the technical efforts related to the development, verification, deployment, operations, and support of the service along with user training in pharmacometrics; (b) the definition and management of the system configuration for delivering analysis results; (c) the translation of the system definition into data management and analysis work breakdown structures; and (d) development of information about productivity for management decision making (2–4).

The application of systems engineering to complex knowledge-generating industries, such as the aerospace and defense industries, has transformed their strategic, tactical, and logistical milieu. The special areas of emphasis embedded in a systems engineering approach have enabled this transformation. These include, first, a top–down approach to design that views the system as a whole, providing a necessary overview and understanding of how individual components and subprocesses effectively fit together; second, a life-cycle orientation that addresses all phases of system engineering, including system design and development, deployment, operation, maintenance, and support; third, a comprehensive effort regarding the definition of system requirements and the traceability of these requirements from the system level downward to specific subprocesses needs; and lastly, an interdisciplinary approach that is used throughout the system design and development process to ensure that all design objectives are addressed in an effective and efficient manner.

36.1.2   Mission of Pharmacometrics

Pharmacometric analysis enables synthesis of disjoint data into knowledge that can inform important clinical development program decisions. Via its modeling and simulation techniques, it is not only a major producer of new knowledge concerning

drug safety and efficacy; it can and should play an integrating role in the entire drug realization process. Thus, the primary mission of every pharmacometrics service organization is to efficiently develop, effectively disseminate, and reliably maintain verifiably accurate and complete explications of determinants of drug effects that cognizant stakeholders can use to enhance the effectiveness of their decisions and improve the efficiency and reliability of their decision-making processes.

A process that will enable the fulfillment of this mission will require the development and deployment of three critical elements. The first is the infrastructure—the data definitions and assembly processes that will allow efficient pooling of data across trials and development programs. The second is the process itself—developing guidelines for deciding when and where modeling and simulation should be applied and criteria for assessing performance and impact. The third element concerns the organization and culture—the establishment of truly integrated, multidisciplinary and multiorganizational development teams trained in the use of modeling and simulation in decision making. Creating these capabilities, infrastructure, and incentives is critical to realizing the full value of modeling and simulation in drug development.

36.1.3 Why Systems Engineering?

In March 2004, the US Food and Drug Administration issued a white paper on the need to improve the system for developing and regulating drugs and medical products (5). This document reflected concern by regulators and the research community about the slow progress in recent years in turning biomedical discoveries into beneficial medical products. The FDA acknowledged that many processes on the "critical path" of product development are inefficient, redundant, and costly. This white paper aimed to stimulate new thinking about how to change and improve current approaches to biomedical product development.

There are many reasons given for failed clinical trials, including flawed clinical trial design, failure to use more powerful study designs, lack of understanding of dose–response, failure to study a broad range of doses, failure to test correct doses, and failure to evaluate Phase 2 results exhaustively (6). These failures will not be corrected by merely recognizing the factors that contribute to the current state of affairs, nor will they be corrected by simply mandating more, faster, and better application of the science of pharmacometrics (7) to support the implementation of the "learn–confirm" paradigm (8). A different kind of science is now required—the science of systems engineering.

Current pharmacometrics practices and available techniques for model-based development are heavily constrained by a vision of model-based development as a subsidiary development process in an empirical development paradigm. Many of the symptoms of the less than optimal functioning of pharmacometrics in this setting are the result of a failure to appreciate the true needs and requirements of the pharmacometrics group. In this chapter, we use the term "pharmacometrics enterprise" to denote the need to transform the current pharmacometrics analysis process to an interdependent enterprise capable of managing the growing complexity of the critical upstream and downstream implications of a fully functional and accountable service. The design of a pharmacometrics enterprise that embodies the required elements is a novel and complex endeavor that can overtax the experience and

intuitive skills of even the most gifted of practitioners. The essential role of systems engineering techniques is to provide practitioners with tools to augment their intuitive experience-based skills and to inform the necessary design decisions.

36.1.4 Chapter Overview

The major theme of this chapter is that any pharmacometrics enterprise capable of executing the mission statement described above must be needs driven not preference, process, product, or service driven. This is not only a major shift in how pharmacometrics services need to be delivered but also a major change in culture in the entire industry. This chapter is designed to fill the void that exists in trying to envision and establish such a pharmacometrics enterprise while meeting quality, cost, and schedule requirements. The chapter is focused on needs and process definitions and not on the technical details of performing modeling and simulations (covered in earlier chapters). The chapter begins with an overview of the general enterprise processes in which a pharmacometrics enterprise must function. Once these higher-level organizational elements are described, we describe the various subprocesses required for a fully functional pharmacometrics enterprise itself. The final section details the prerequisite architectural traits required for effectively deploying a fully integrated pharmacometrics enterprise in an existing drug development enterprise, including structural changes in the parent or client enterprise processes along with additional informatics support.

36.2 ENTERPRISE OF PHARMACOMETRIC ANALYSIS

Like any other enterprise, a pharmacometrics enterprise exists to accomplish the purposes of its stakeholders. It is important to recognize that "stakeholder" in the context of a pharmacometrics service refers not only to the parent organization in the case of an internal pharmacometrics group or client in the case of an external consulting service, but also to the myriad "customers" of the information products provided by a pharmacometrics service. These external customers certainly include regulatory agencies, but also health care delivery providers, such as managed care organizations, physicians, pharmacists, nurses, pharmacy and therapeutics committees, the scientific community, and patients.

The effectiveness of a pharmacometrics service—and its fundamental value to stakeholders—ultimately consists in the degrees to which the outcomes of its actions correspond to their purposes. The ability of an enterprise to achieve and reliably maintain this purpose rests, in turn, on the effectiveness of its actions—that is, how capable the enterprise is and how effectively it deploys its capabilities. Pharmacometrics service capability consists in the degrees and extents to which the quantitative characterizations of the determinants of drug effects it delivers satisfy the informatic needs of its customers. Deployment effectiveness for a pharmacometrics service consists in its ability to consistently deliver analytic services within cost, schedule, and quality specifications.

The preceding is meant to emphasize that the capability of a pharmacometrics enterprise is directly determined by the scope, effectiveness, and efficiencies of the processes that the enterprise executes and the infrastructure that supports those

processes. That is, establishing a capable pharmacometrics enterprise—or, for that matter, reengineering an existing one—consists in designing (or redesigning) and implementing (or reimplementing) these processes and infrastructure, so that the enterprise can (a) accurately ascertain, be principally driven by, and consistently sustain its focus on the actual needs of its stakeholders; (b) reliably deliver solutions to those needs, and (c) be efficient at both of the previous points.

Having arrived at this conclusion, the obvious questions we need to answer in order to establish this enterprise are the following:

1. What are the processes of a pharmacometrics enterprise?
2. What specific infrastructure is required to support those processes?
3. What criteria should be used to assess designs and implementations of those processes?

We address general elements of the answer to the first question in the remainder of this section. Its pharmacometrics service-specific elements are presented in Section 36.3. Answers to the second and third questions are provided in Section 36.4.

36.2.1 Generic Enterprise Processes

All enterprises share many architectural features and characteristics. For our purposes here, we describe the three essential classes of *generic* enterprise processes and, by doing so, get the first part of a general answer to the process question posed above. As depicted in Figure 36.1, the three most important of these are the realization processes comprising the complete life cycles of three classes of entities: strategy, agency, and solution.

1. *Strategy.* Strategy realization is the enterprise's process for determining what the actual needs implicit in and antecedent to the purposes of its stakeholders are. This process translates information and intelligence, including that about

FIGURE 36.1 Generic enterprise process matrix.

stakeholder purposes, into concrete objectives, programs, and directives for governing the enterprise.

2. *Agency.* In order to execute a strategy, an enterprise must have some process or means (agencies) for addressing the systemic infrastructure needs required for the pursuit of its mission. This is referred to as the agency realization process. Three familiar examples of pervasive agencies are information, human assets, and matériel.

3. *Solution.* The output of an enterprise is generically referred to as a solution and it can be a product, a service, or, in the case of a pharmacometrics enterprise, information about the determinants of drug safety and efficacy. Solution realization is a general term for processes that define and deliver solutions to the needs and requirements that stem from stakeholder purposes. The pharmacometric analysis process would be an example of a pharmacometric solution realization and it is described in detail in Section 36.3.

All processes share certain structural characteristics, and the second part of our general answer to the process question before us can be obtained by exploiting that fact. All "end-to-end" or "total life-cycle" realization processes are subdivided into three distinct classes of realization *subprocesses* we call *definition*, *development*, and *execution*.

1. *Definitional Realization Subprocesses.* Definitional subprocesses are initialization processes. These processes transform preliminary realization process inputs, such as stakeholder purposes, systemic requirements, and technical needs, into prescriptive characterizations of objectives, capabilities and infrastructures, and both internal and external products and services.

2. *Developmental Realization Subprocesses.* Developmental subprocesses are provisioning processes. These processes translate formal definitions of objectives, enterprise infrastructures, and products (services) output by definitional subprocesses into operational programs, infrastructures, and informational systems.

3. *Execution Realization Subprocesses.* Execution subprocesses are operationalization processes. These processes invoke the programs and systems produced by developmental subprocesses. These processes include (a) governance, the subprocess that initiates action programs and guides enterprise operations; (b) operation, the enterprise realization subprocess that invokes enterprise action programs, yielding the operations that carry out enterprise missions; and (c) application, the solution realization subprocess that invokes technical systems in actionable contexts, thereby yielding instrumental solutions to stakeholder needs and requirements.

Having now obtained both "parts" of the general answer to the process question, we can "assemble" them to get the "whole" generic answer. We do this by combining the first threefold entity/process factorization described initially with the second threefold phase/process factorization just sketched. To visualize this we assign the REALIZATION subprocess classes to the "X" axis and the processes associated with the three entity types to the "Y" axis, giving us the matrix structure depicted in

Figure 36.1. The elements of this matrix depict the nine main generic processes of any needs-driven solution-oriented enterprise.

36.2.2 Specific Pharmacometrics Service Enterprise Processes

A pharmacometrics enterprise is, intrinsically and by definition, a needs-driven solution-oriented enterprise. Therefore, establishing a pharmacometrics service consists of designing and implementing specific *variants* of the nine processes enumerated above and the infrastructure required to execute and sustain them. The core process of a pharmacometrics service enterprise is pharmacometric analysis, conceived of and visualized in Figure 36.2 as a specific variant of SOLUTION REALIZATION.

Developing a pharmacometrics service enterprise fundamentally consists in iteratively accomplishing three major tasks. The first task is to design and operationalize the core process. We describe it in Section 36.3 and provide guidelines for designing and implementing it in Section 36.4. The second task is interface design and implementation. The process of pharmacometric analysis is invoked for various reasons and in various phases of the drug realization process, including drug discovery, development, and commercialization, to fill gaps in knowledge of relationships between determinants of drug effects and those effects. Installing and integrating a pharmacometrics enterprise into a drug development enterprise entails designing and implementing the interaction protocols that define the interfaces between the pharmacometrics enterprise and the entire drug realization process. One of the challenges in building and maintaining an effective pharmacometrics enterprise is

FIGURE 36.2 Pharmacometric analysis as a SOLUTION REALIZATION variant.

the fact that these interfaces have not been well defined from the perspective of the pharmacometrician nor the drug realization process itself. The third task is the design and implementation of the supervening processes of pharmacometric analysis depicted in Figure 36.2. It should be emphasized that a pharmacometrics enterprise encompasses a great deal more than just the scientific and technical processes for performing analyses and for communicating the results typically considered the core activities of a pharmacometrics service. Establishing a pharmacometrics enterprise also requires attention to less visible but nonetheless crucial supervening elements of a total pharmacometrics solution.

As drug development organizations (and their regulatory stakeholders) seek to improve effectiveness and performance of their core discovery, development, and evaluation processes, the role of pharmacometrics in drug development is rapidly shifting from a sideline discipline to a mainstream function. This shift, in turn, is generating expectations that pharmacometrics service organizations should assume larger and more critical roles in these processes. A systematic, requirements-driven approach is essential for meeting these new architectural—as opposed to technical—challenges, and since stakeholder needs constitute the ultimate sources of all valid enterprise requirements, some process for systematically identifying needs and requirements and for deriving the architectural structures and features is required.

Three of the supervening enterprise processes highlighted in Figure 36.2 are particularly important in the context of engineering or reengineering a pharmacometrics enterprise in an industry undergoing a profound transformation. These three key processes include Intent Formalization, Enterprise Definition, and Enterprise Programming.

36.2.2.1 *Intent Formalization*

Intent formalization is the definitional subprocess for translating stakeholder intents and measures of acceptability (MOA) into actionable mission definitions, goals, and strategic objectives (9). This process is a critical one for pharmacometrics services for two specific reasons. First, *some* process must exist for defining actionable objectives and for ensuring that those objectives constitute the best possible balance of stakeholder intents. Second, and more importantly from a mid- to long-term strategic perspective, both the pharmacometrics enterprise and the drug realization process are undergoing a remarkable structural transformation, and intent formalization is a critical element of the larger process for managing this architectural change. This will be addressed in more detail in Section 36.4.

36.2.2.2 *Enterprise Definition*

The enterprise definition subprocess translates the agential (process, organization, information, and infrastructure) needs and requirements deriving from mission definitions into enterprise architectures—that is, into designs for enterprise operating systems capable of executing those missions (2). The design of a pharmacometrics service capable of fulfilling its mission is inextricably linked to the enterprise architecture of the drug realization process deployed by the parent organization. This link will become more apparent—and urgent—as the transformation to model-based development becomes further advanced. The relationships between the pharmacometrics process and the drug realization process are more closely intertwined than

is evident in the current empiric-based development paradigm and the emerging difficulties with meeting expectations within cost, quality, and schedule constraints will become ever more obvious as we move toward model-based development.

36.2.2.3   Enterprise Programming

Enterprise programming is the developmental subprocess for translating mission definitions and objectives into executable action programs for carrying them out (4). Designing and implementing this process is also critical for a pharmacometrics service for two specific reasons. First, *some* process must exist for defining an actionable and executable program and for ensuring that the program is properly designed from a business perspective; for example, that it is financially sustainable. Second, and more importantly from a mid- to long-term operational perspective, the pharmacometrics enterprise design must continually evolve to effectively address the implications of structural changes that are occurring at all levels of the pharmaceutical industry.

36.3   PROCESS OF PHARMACOMETRIC ANALYSIS

Pharmacometric analysis is the enterprise-specific SOLUTION REALIZATION process for a pharmacometrics enterprise and is, therefore, a core enterprise process. Designing an implementation of this process, setting up an organization to execute and manage it, and building an infrastructure to sustain it are, accordingly, the core tasks involved in establishing a pharmacometrics enterprise, or in undertaking to improve an existing one.

Integrating both pharmacokinetic and pharmacodynamic analyses, a life-cycle complete pharmacometric analysis process comprises seven major subprocesses, as depicted in Figure 36.3. That is, any comprehensive pharmacometrics enterprise will be capable of executing and governing the complete process—all seven major subprocesses—from invocation to termination. Of course, not every performance of the service will necessarily require the comprehensive and complete pharmacometrics enterprise.

Pharmacometric analysis is invoked by the formulation of queries to address gaps in knowledge of relationships between determinants of drug effects and those effects. Its principal inputs are target product profiles defining pharmacological and pharmacoeconomic baselines, designs and protocols for clinical trials aimed at demonstrating safety and efficacy in subject populations, and prior knowledge in the form of information bases and technical literature. The principal outputs of pharmacometric analysis are responses to queries constituting quantitative explications of determinants of drug effects. These typically can include characterizations $(D \rightarrow C \rightarrow E)$ of drug kinetics, drug dynamics (and under these, characterizations of various pharmacological magnitudes determining safety and efficacy attributes), and predictions or extrapolations concerning any of the preceding.

Illustrated by the schematic pictured in Figure 36.4, the subprocess elements we are concerned with in the descriptions that follow are focused on the architecture of the process, not its operation in a specific development program. The focus here is to provide a basis for engineering or reengineering a service, not to describe the performance of the service once it is deployed.

FIGURE 36.3 Pharmacometric analysis process elements.

FIGURE 36.4 Pharmacometric analysis subprocess element scheme.

36.3.1 Analysis Planning

The first step required to address a knowledge gap is to identify and characterize it. After all, even the most cogent answer to the wrong question is useless, and an incomplete answer—even to the right question—is frequently only marginally less

so. Hence, the quality of the final product of an analytic effort is directly and heavily dependent on the effectiveness of the plan that was executed to produce it; and in the case of a pharmacometric analysis, how precisely identified and how completely characterized the knowledge gap was to begin with.

The second step required to address a knowledge gap is to develop an approach to resolve it. Here one will inevitably confront the additional complexity of alternatives. Presented with options, devising an analysis approach entails ascertaining which among them offers the best balance between analytic effectiveness on one hand, and executability in terms of capability, cost, and schedule requirements on the other. After all, no matter how scientifically potent and technically effective, an approach that cannot be executed is worthless, and one that is unnecessarily difficult and needlessly inefficient will likely be only marginally better.

The third step required to address a knowledge gap is to devise contingencies for potential outcomes of the pharmacometric analysis. These include (a) the potential that the data collected are noninformative, (b) the generation of unanticipated results that are contrary to the expected outcome, and (c) the generation of results that should have a significant impact on the subsequent design of the development program. Each of these potential outcomes has important and iterative effects on the analysis plan itself and the subsequent modeling and simulation activities.

36.3.2 Data Assembly

The assembly of data sets for a pharmacometric analysis is complicated by the complexity of the content, the origin of specific data elements, and the structure of the required database. These analyses typically require pooling disparate data, including drug concentrations, drug dosing histories, patient demography, laboratory data, use of concomitant medicines, and measures of efficacy and safety to create a time-ordered sequence of events for each patient from the time of enrollment in a trial until its conclusion. This information must be assembled from numerous databases often managed by different departments, either internal or external to the company. As a result, data assembly can be a cumbersome and time-consuming process. There has been considerable effort to create data definition standards to facilitate data assembly; however, these efforts are focused on empirically based development programs (10,11). It is not our intent to review well known points about data management practice and related quality management issues, but instead to emphasize that there are important semantic issues that are not typically addressed by these efforts but that nevertheless have important impacts on data assembly for pharmacometrics services. The strategy for addressing semantic issues in the standard data management paradigm is wholly inadequate for the time-dependent and sequential individual histories required for a typical pharmacometric analysis.

The challenges of data assembly go beyond these semantic issues, however, and include process-related problems involving data scrubbing and quality assurance. A number of essential data queries, particularly the determination of whether the drug concentration values and the date and time of sampling make sense in the context of the dosing history, cannot be performed until the drug concentrations have been merged with the drug dosing history for each patient. Yet it is common for these individual data elements to be queried separately, so the important questions as to whether data issues will impact on the quality of the results or preclude any

meaningful analysis may not be recognized in a timely manner (12). Moreover, recording errors in sampling and dosing time, incomplete data collection, and administrative errors noted after merging the drug concentration and dosing histories can result in the loss of a substantial amount of valuable and expensive data, underscoring the need for pharmacometric-oriented monitoring in clinical trials incorporating pharmacometric sampling. Ideally, pharmacokinetic data should be queried during trial execution to identify problems early so that they can be rectified by appropriate interventions at the problem sites (13).

36.3.3   Exploratory Analysis

While the pharmacometric analysis plan combines one part wishful thinking and two parts prior or emerging knowledge of a drug's pharmacokinetics and pharmacodynamics, the exploratory data analysis process serves as a rapid prototyping exercise to determine what can be reasonably asked of the data set. The effectiveness of the exploratory analysis process centers on assessing the confirmatory power of the data set to provide accurate, reliable, and relevant results from model development. The efficacy of the exploratory analysis process concerns the degree to which it highlights issues, shortfalls, strengths, and weaknesses of the analysis plan.

The exploratory data analysis process must address issues that arise from a number of perspectives. Typically, the modeler will develop a series of graphic and tabular displays to explore the data and determine the quality and quantity of the data. Most pharmacometrics services have developed a standard list that is supplemented as needed to address the special characteristics of a particular development program. These graphs represent an inexpensive approach to identifying gaps in the data that may have arisen because of performance problems in the clinic where the data were originally collected, such as an inability to follow the proscribed sampling strategy because of clinical realities or miscommunications, and errors in data collection or reporting. These problems are exacerbated if the study monitors do not incorporate checks for pharmacometric data into their monitoring activities during site visits. Consequently, errors may not be identified until the analysis-ready data sets are created and examined. Importantly, knowledge about the pharmacology of the drug rapidly evolves during the early stages of drug development, particularly as the studies move from healthy volunteers to patients with various degrees of severity of illness. Consequently, modeling objectives and feasibility may change from the time the study protocol was developed to the time data are available for analysis.

36.3.4   Model Development

Obviously, one of the key tasks of a pharmacometrics service is the development of pharmacokinetic and pharmacodynamic models that serve as a mathematical representation of physiologic or pharmacologic phenomena. This is also one of the main activities that create a mockery of cost and schedule estimates. The complexity and scope of a pharmacometric model are limited only by the imagination, literature access, and computer resources of the pharmacometrician. Consequently, an effective model development process is one in which the extent of achievable

correspondence between the model and reality is balanced by time constraints and limitations of the available data. Efficiency in this case is the interpretability and applicability of the model in the context of the objectives of the analysis. This latter point is critical to adhering to cost and schedule restrictions. Model development is by definition a learning and confirming activity. The results of exploratory analyses may shed light on the properties of the phenomena that must be incorporated into the model, but model development activities themselves may allow other heretofore unappreciated complexities to emerge.

It is critical to achieve a balance between respect for timelines—timelines that are becoming increasingly compressed as pharmacometricians are required to meet the same schedules for modeling and simulation as statisticians for traditional statistical analyses—and respect for the learning process that unfolds as new knowledge is gleaned from the model-building exercise. Without this balance, we model until the time runs out and then whatever is available is good enough. It is far better to focus on the critical path of model development, leaving ancillary issues to be further developed later as time permits, than it is to allow the modeling project to be sidetracked by those ancillary issues with the subsequent loss of time and ability to attend to critical path issues. It is essential for the pharmacometrics service to develop measures of acceptability that can guide the model-building process and serve as indicators for monitoring modeling progress and distinguishing between critical path and ancillary issues.

In a fully realized model-based development paradigm, models will be both the instruments and aims of drug development programs (14). There is a much more intimate relationship between the premises, hypotheses, and theorems that a model realizes or conveys, on the one hand, and those that a clinical trial is meant to test. In other words, the model-based paradigm will focus on the development and support of models as the primary outcome of a development program. This entails a much more iterative process than is currently employed and requires a more rigorous and efficient "enterprise engineered" process to support timely decision making.

36.3.5 Simulation

Simulation has emerged as an important tool for extrapolating from scenarios that generated the data for model development activities into scenarios of potential interest for the drug development program (15). Similar to the model development process, effectiveness in the case of simulations is the correspondence between the model and reality. Efficiency hinges on the applicability of the results as the target scenarios step outside the boundaries of the initial model. Here we need to ask about the extensibility of the simulation results and how broadly applicable the results are.

Simulations have been promoted as a powerful method for predicting the outcomes of various drug development scenarios and hence for designing more cost-effective development programs. There is a dependency between the available knowledge and data for model development and the subsequent extension of the model via simulation that influences both directly and indirectly the extensibility of the process. Pharmacometricians lack the tools for assessing the gaps in knowledge that impact on this extensibility, and in fact there are little data in the public domain

with respect to previous experiences with the congruence between study outcomes predicted by simulations and the actual clinical trial results. Thus, the true impact of the widening gap as we move from data to model to simulation predictions is unknown. Importantly, without these data, we lack an essential feedback loop for determining strategies for improving the reliability and robustness of clinical trial simulations.

36.3.6 Validation

Model validation can be defined as the process of substantiating that the model within its domain of application provides the required functionality, including input and output variables, and that the values it computes are sufficiently accurate for the intended use (16). This requires investigations of both functionality and accuracy requiring criteria for judging acceptability that can be used to determine if the model is useful and appropriate. Investigation of the structure of a model requires, for example, studying whether the relationships and assumptions are based on general accepted theory and whether all variables considered relevant have been taken into account. Structure-oriented behavior tests assess the validity of the model structure indirectly by running the model. One method is to study the behavior of the model by entering extreme input values. Another approach involves a sensitivity analysis entailing an analysis of the model assumptions and of the influence of plausible variations in parameters, structure, and possible exogenous variables. The sensitivity of the outputs in response to changes in parameters and structure can be studied to see whether these are realistic. Data splitting and external validation entails a comparison of the model to the system that has been modeled using data that have not been used for model construction. Finally, the usability of the model should also be investigated. The model should be geared to problems and questions that are thought to be important by relevant stakeholders. Model validation is technically defined to take place after model development, but this process should take place in every stage of the analysis process. Although model validation is sometimes conducted as a separate exercise from model development, validation is key to building confidence in the model.

36.3.7 Presentation

The principal stakeholders in the drug development process, particularly the development team, the regulatory review team, and the marketing group, have come to expect and make decisions on a binary outcome of a clinical trial as provided by the "*p*-value" driven concept of efficacy and safety (1). The outcome of a pharmacometric analysis is considerably more complex because it describes a continuum of outcomes, and decisions are based on setting threshold values that have important downstream implications. This complexity can be compounded if the teams are subjected to a debate among the members of the pharmacometrics group as to the relative value of a two- versus a three-compartment model in improving the goodness-of-fit plots. Pharmacometricians must recognize that the main challenge of the development team is to integrate emerging information on safety and efficacy into decisions on whether and how to continue the development program. The strategic importance of pharmacometric analyses in this decision-making process

must be emphasized in presentations to the development team. The pharmacometrician's scientific assessment as to the underlying appropriateness of the model is best challenged and defended in presentations to peers in the pharmacometrics group, where the emphasis is on the quality of the fit and the appropriateness of decisions made during the model-fitting process.

Considerable effort is required for the preparation of technical reports, specialized presentation to development teams, regulatory interactions, and publications and presentations (see also Chapter 37). The efforts required to produce these materials are frequently duplicative and may at best involve repurposing content, requiring considerable rework and editing to achieve the desired content, only to realize that specific information is not readily available or was not previously generated during an appropriate opportunity in the pharmacometrics process. This issue speaks to important deficiencies in the representational infrastructure of pharmacometrics and is addressed further in Section 36.4.2.

36.4 PHARMACOMETRICS ENTERPRISE DESIGN

Now that the components of the pharmacometrics service process have been defined, we can return to enterprise design issues and further develop the implications of the interfaces to drug realization processes and informatic infrastructure. The enterprise architecture described in Section 36.2 is of critical relevance to the successful design or reengineering of a pharmacometrics service, whether it is an internal group embedded in a larger drug realization enterprise or a stand-alone enterprise providing services on an outsourcing basis. There are two important reasons why the entire enterprise, and not just the process of performing analyses, must be kept in mind. First, the broad-based impact that pharmacometrics services are having on development and regulatory decision making highlights the need to interface with a number of departments and processes across the drug realization process. Second, the queries that invoke a pharmacometric analysis originate for varying reasons during various stages of drug discovery, development, and commercialization. Figure 36.5 depicts the capability matrix for the drug realization process and shows the myriad potential interrelationships to the pharmacometrics enterprise.

The effectiveness of a pharmacometrics service ultimately consists in the degree to which the results of modeling and simulation activities correspond to the expectations of the stakeholders, including those requesting services and those who will be the recipients of work products. The ability to reliably meet expectations rests on how capable the pharmacometrics service is and how effectively it deploys its capabilities. Reliably determining the requisite degrees of correspondences between capabilities and expectations is a difficult yet nevertheless crucial process for sustainable success that must draw on strategic management and systems engineering techniques.

From a strategic management perspective, there are three enduring strategic challenges that must be successfully addressed:

1. Organizational alignment—ensuring that the pharmacometric analysis enterprise will be focused on the right things.

FIGURE 36.5 Capability matrix of the drug realization enterprise.

2. Organizational effectiveness—ensuring that the pharmacometric analysis enterprise will be capable of delivering the right things.
3. Organizational performance—ensuring that the enterprise will be efficient at addressing both these challenges.

From a systems engineering perspective, there are three enduring technical challenges that any pharmacometrics service enterprise must successfully address:

1. Service applicability—ensuring that pharmacometric analysis findings and communications are suited to the purposes of its stakeholders.
2. Service effectiveness—ensuring that these offerings will reliably facilitate capable accomplishment of those purposes.
3. Service performance—ensuring that their use engenders efficiency in both these dimensions.

Rigorous techniques for technical effectiveness and performance measurement exist and have been used for years and must be mined for applicable lessons and directions for the pharmacometrics enterprise. The development of an integrated solution to these strategic and technical challenges is a particularly difficult, but nonetheless crucial, effort that will define the level of performance achievable by a pharmacometrics service in model-based development.

36.4.1   Essential Pharmacometrics Service-Specific Traits

There are four key requirements that, programmatically speaking, constitute critical success factors for the desired level of effectiveness of the pharmacometrics enterprise. These requirements stem from the shift in roles of the pharmacometrics enterprise that mandates a clear formalization of intents from the larger drug realization enterprise and provide answers to several key questions: What are the actual requirements for the hosting organization? What are the implications of deploying a solution to those requirements? What resources are available and would be required to pursue such an initiative?

1. *Definitive Characterizations of Stakeholder Aims.* The purposes of a pharmacometrics enterprise must be explicitly defined to enable action programs, with their upstream and downstream implications for the drug realization process, to be devised and executed. Purposes typically exemplify tremendous variations in several dimensions. They range from the abstract to concrete; from the unattainable to the trivial; from the short-term to the long-term. Some purposes are identical, some are congruent, and some are diametrically opposed. Some purposes depend on the successful attainment of others in order for them to be accomplished.

The capabilities and resources of even the most well funded pharmacometrics enterprise are finite, and it is almost always impossible to achieve all of the aims of its stakeholders to the degrees they desire. An achievable and acceptable balance must be struck. Requirements antecedent to the purposes of the pharmacometrics enterprise must be identified to enable the synthesis, development, and use of specific solutions for implementing the analysis process. Qualification criteria must be explicitly defined to enable governance of the actions of pharmacometricians and to enable determination of the value of the solution.

2. *Demonstrable Traceability.* Proposed pharmacometric enterprise capabilities and assets must constitute demonstrable solutions to the strategic, technical, and infrastructure requirements entailed by stakeholder aims. This in turn mandates an enterprise design, or architecture, that is provably traceable to the mission requirements. This enterprise architecture requires the identification of the needs and requirements stemming from mission definitions, synthesis of system element designs entailed by the needs and requirements, and an evaluation process based on measures of acceptability.

3. *Executable Operations.* Accomplishing stakeholder intents requires effective action on the part of the pharmacometrics enterprise. Effective action presupposes prior definition and implementation of specific action programs. Action program definition must be governed by additional constraints over and above the functional, performance, and effectiveness requirements represented by enterprise objectives and measures of acceptability. One of the most crucial of these is *executability*. The mission definition for the pharmacometrics enterprise, the capability assessments of the agencies to be deployed, coupled with action definitions, operating plans, and resource allocations must all be defined and aligned to ensure that the enterprise is actually capable of executing its mission.

4. *Actionable Assurance Process.* The results of analyses performed by the pharmacometrics enterprise will not always align with stakeholder expectations. In fact, expectations may themselves be invalidated by changes in the environment, includ-

ing emerging knowledge from the development program and shifts in regulatory policy. A process for continuously determining the degrees to which the outcomes of pharmacometric enterprise actions correlate to stakeholder aims is essential for success, as is a process for effecting changes when they do not.

36.4.2 Pharmacometrics Service-Specific Design Criteria

The challenges of designing a pharmacometrics enterprise that can achieve requisite levels of efficiency, effectiveness, and capability stem from the unmet, and in some cases poorly recognized, needs of the pharmacometrics analysis process. Many of these needs can be addressed by using the following criteria to guide enterprise engineering efforts.

36.4.2.1 *Maximize Visibility*

Because the predominant drug development paradigm is based on empirical methods, there is a need to more clearly and convincingly convey the pharmacometrics value proposition. The model-based development paradigm enabled by pharmacometric analysis is very different from the empiric-based paradigm. As a result, the value of these services to development programs and their stakeholders is frequently difficult to identify and readily communicate. This inability to definitely demonstrate strategic and economic value of model-centric development services in empirical development contexts constitutes a clear challenge. As noted earlier, there can be a lack of familiarity with complex characterizations of safety and efficacy among members of the development teams. Well crafted presentations of results focused on the decision-making implications of modeling and simulation results are key to promoting the visibility of the service and the importance of the findings (see Chapter 37). These communications, at critical junctures of the pharmacometric process, provide (a) an opportunity to evaluate the value of the emerging knowledge; a reliable and dependable method for (b) identifying and explaining modeling issues and managing expectations with respect to challenges stemming from flaws in study design, data collection, or modeling limitations; and (c) teaching opportunities with respect to the needs and requirements for future modeling and simulation applications.

36.4.2.2 *Maximize Multilateral Action Potential*

There are two fundamental ways an enterprise can execute a process. The first is unilaterally—the enterprise acts alone or at least is the dominant execution and governance agency in the process. The second is multilaterally—the enterprise acts in concert with others. Unilateral action is efficient but requires total command of all the resources and capabilities to execute and govern a process. Multilateral action is less efficient because it requires consensus building and coordination. However, multilateral action does not require total command of all the resources and capabilities to execute a process; it is a strategy to accomplish stakeholder purposes with less. The intensely competitive and global pharmaceutical business environment places severe constraints on the resources available to a pharmacometrics enterprise to achieve and sustain success. The capability to act in multilateral partnerships with various stakeholders is a very effective means to achieve this success. Thus, the design of an effective pharmacometrics service should be assessed in terms of how

it will augment the capability of the enterprise to act multilaterally or to participate in a multilateral process.

36.4.2.3 Skill Augmentation

The analytical process employed by a pharmacometrics service is growing increasingly sophisticated, and customers both upstream and downstream of the service are placing an increased emphasis on the results as a basis for drug development and regulatory decision making. The ability to perform these analyses with the requisite level of timeliness, quality, and sophistication creates the need for providing teams with personnel with appropriate levels of scientific, technical, and business skills. Much has been said about the need for skilled pharmacometricians, but the challenge is larger than that. A career path that provides competent scientists with the ability to acquire the increasingly sophisticated skills to move from practitioner to strategist is essential. The need for this latter individual, with the ability to recognize when and how modeling and simulations will be truly cost effective, is emerging now. The training for these scientists/architects must encompass not only the technical and scientific aspects of pharmacometric analysis but also the skills required for assessing and balancing risk, performance, schedule, and cost considerations. These scientists/architects will be key to the successful implementation of multilateral processes increasingly being fielded by Pharma, such as the outsourcing of pharmacometric analyses.

36.4.2.4 Eliminate Unnecessary Variability

In the process of embarking on a pharmacometric analysis, a scientist is confronted with a number of different sources of diversity. There is the diversity that stems from differences in expected drug effects across a therapeutic class, differences in strategies for analysis and presentation of modeling and simulation results, and differences in process execution including variability due to contingent events and pharmacometrician capabilities and preferences. These sources of variability are commonly run together with the unfortunate effect of making each pharmacometric analysis appear to be a "one-off" or a unique creation. By allowing the sources of diversity to remain unbridled, pharmacometricians will continue to experience efficiency shortfalls with attendant effects on the cost, schedule, and quality attributes of analyses. Of course, an uninformed alternative is to implement a process that mandates conformity with subsequent loss of flexibility and responsiveness.

One must not confuse the challenges of performing a pharmacometric analysis in an inherently variable environment with the problems of elucidating the nature of novel drug effects. Pharmacometricians are very familiar with a number of dimensions of variability acting as sources of requirements for pharmacometric analyses. These include its ADME properties, its phase of development, whether it is a new entity or a new member of an existing class, whether it employs a new or existing biomarker, and whether it is a small molecule or a protein. While these sources of variability are inherent in the process, other sources of variation are amenable to systemization.

Pharmacometric analyses performed in support of drug development decision making require a diverse variety of models and strategies for using the results of simulations based on those models to inform decision making during all phases of drug development. Understanding what these models need to look like entails a

clear understanding of the requirements of the drug development processes these modeling and simulation activities are attempting to satisfy (17). Pharmacometric models are superficially, and to some extent morphologically, similar from project to project. Moreover, if one looks carefully at the elements of a pharmacometrics report, one will find a very large number of highly similar elements and a significant number of practically identical ones. Pharmacometricians must seek to minimize the effects of variability in the modeling process and eliminate to the greatest extent possible unnecessary variability in work processes and products in order to realize efficiency gains with attendant effects on the cost, schedule, and quality attributes of analyses.

36.4.2.5   *Strengthen Representational Infrastructure*

Data management problems and report production and configuration management problems are converse sides of the same coin. They are symptomatic of two distinct sets of representational capability gaps. The first consists in definition data management shortfalls—so called schematic gaps. The second set consists in implementation deficiencies—shortcomings in our existing software systems.

The knowledge content encompassed by final technical reports and associated presentations should be understood ontogenetically, and there should be specific emphasis on the milestones to delimit key transitions in the knowledge generation. Performing an analysis of final technical report elements and driving all systemization activities of the representational infrastructure backward from this analysis can best address the nontechnical, informatic needs of a pharmacometrics enterprise. This schematization would enable the ability to go from analysis plan to final technical report by adding in elements derived from the pharmacometrics process, as they are available, and reduce the considerable time and effort typically involved in producing and maintaining technical reports.

In addition, some of the challenges in meeting cost, quality, and schedule requirements stem from archaic work processes. For example, capturing the requirements for a project in a new therapeutic area can be difficult because of time constraints and unfamiliarity with the knowledge content. The therapeutic area may be unfamiliar to scientists and support staff, in which case literature searches, article analysis, and communication of relevant elements of those to support staff are necessary precursors to formulating a cogent analysis plan. Systematizing the knowledge needed for a new project via comprehensive taxonomic analysis of the pharmacometric relations used to explicate these relationships will be critical to either circumventing or dramatically reducing the literature search/analysis/communication cycle.

36.5   SUMMARY

The pharmaceutical industry is undergoing major structural change in several dimensions simultaneously, and these changes will completely transform the industry's fundamental business models, core processes, and socio-techno-logistical infrastructures. The accelerating shift from empirical to formal (i.e., model-based) methods and the growing reliance on modeling and simulation in decision making are forcing important and urgent changes to occur in the nature of the work of a pharmacometrics enterprise. Accordingly, what it takes to establish a pharma-

cometrics enterprise (or sustain and strengthen an existing one) is also changing significantly from what it has required in the past.

Effectively meeting the challenges one confronts in establishing a pharmacometrics enterprise in the face of this shift in the developmental center of gravity mandates the use of systematic and rigorous methods to consciously and deliberately "engineer" the service, or "reengineer" it, as the case may be. That is, successfully deploying a complex technical service such as pharmacometric analysis in today's complex and fluid environment mandates prior development of robust enterprise designs and effective implementation programs. Doing this right, in turn, requires use of a process that is designed explicitly to do this sort of job—the systems engineering process. The essential role of systems engineering techniques is to provision practitioners with tools to augment their intuitive experience-based skills and to inform the decisions they must make in the design of a pharmacometrics enterprise.

Pharmacometrics, in its current state within the drug realization process, is faced with a significant, but ephemeral, opportunity. The tools of pharmacometric analysis are sufficiently understood at the same time that the limitations of empiric-based development are becoming more widely appreciated. However, tools do not an enterprise make and ad hoc solutions incur a high risk of failure in the face of cost, quality, and schedule constraints. Only by envisioning and engineering a complete pharmacometrics enterprise can the full promise of model-based development be realized. The unsatisfactory—and short-lived—alternative is an underperforming pharmacometrics service with a reputation for failing to meet expectations. This can only result in an erosion of support and a surrendering of a singular opportunity to realize the industry and societal implications of model-based development.

ACKNOWLEDGMENTS

REFERENCES

1. T. H. Grasela, J. Fiedler-Kelly, C. A. Walawander, J. S. Owen, B. B. Cirincione, K. E. Reitz, E. A. Ludwig, J. A. Passarell, and C. W. Dement, Challenges in the transition to model-based development. *Themed Issue: Population Pharmacokinetics—In Memory of Louis Sheiner, AAPSJ* [serial online] **7**:E488–E495 (2005). DOI: 10.1208/aapsj070249.

2. B. S. Blanchard and W. J. Fabrycky, *Systems Engineering and Analysis.* 3rd ed. Prentice Hall, Englewood Cliffs, NJ, 1998.

3. A. P. Sage and W. B. Rouse (Eds.), *Handbook of Systems Engineering and Management.* Wiley, Hoboken, NJ, 1999.

4. F. R. David, *Strategic Management Concepts.* Prentice Hall, Englewood Cliffs, NJ, 1999.

5. *Challenge and Opportunity on the Critical Path to New Medical Products.* FDA Department of Health & Human Services, Rockville, MD, March 2004. `http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html`.

6. *Drug Development Science: Obstacles and Opportunities for Collaboration Among Academia, Industry and Government.* Association of American Medical Colleges, FDA Department of Health & Human Services, Rockville, MD, 2005. `https://services.aamc.org/`.

7. D. A. Berry, P. Müller, and A. P. Grieve, *Adaptive Bayesian Designs for Dose-Ranging Trials. Case Studies in Bayesian Statistics*, Vol. V. Springer Verlag, New York, 2002, pp. 99–181.

8. L. B. Sheiner, Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

9. J. E. Armstrong, Issue formulation, in *Handbook of Systems Engineering and Management*, A. P. Sage and W. B. Rouse (Eds.). Wiley, Hoboken, NJ, 1999, pp. 993–996.

10. *Good Clinical Data Management Practices*, Version 3. Society for Clinical Data Management, Inc., 2003. `http://www.scdm.org/GCDMP/`.

11. *Statistical Analysis Dataset Model: Change from Baseline Version 0.2.* Prepared by the CDISC Analysis Dataset Modeling Team (AdaM), 2005. `http://www.cdisc.org/models/adam/V1.0/ADaMChangefromBaselineV0.2.pdf`.

12. E. J. Antal, T. H. Grasela, and R. B. Smith, An evaluation of population pharmacokinetics in therapeutic trials. Part III. Prospective data collection versus retrospective data assembly. *Clin Pharmacol Ther* **46**:552–559 (1989).

13. T. H. Grasela, E. J. Antal, J. Fiedler-Kelly, D. J. Foit, B. Barth, D. W. Knuth, B. J. Carel, and S. R. Cox, An automated drug concentration screening and quality assurance program for clinical trials. *Drug Info J* **33**:273–279 (1999).

14. L. B. Sheiner and J. L. Steimer, Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu Rev Pharmacol Toxicol* **40**:67–95 (2000).

15. N. H. G. Holford, H. C. Kimko, J. P. R. Monteleone, and C. C. Peck, Simulation of clinical trials. *Annu Rev Pharmacol Toxicol* **40**:209–234 (2000).

16. C. E. van Daalen, W. A. H. Thissen, and A. Verbraeck, Methods for the modeling and analysis of alternatives, in *Handbook of Systems Engineering and Management*, A. P. Sage and W. B. Rouse (Eds.). Wiley, Hoboken, NJ, 1999, pp. 1037–1073.

17. P. R. Bonate and D. R. Howard, *Pharmacokinetics in Drug Development, Volume 1: Clinical Study Design and Analysis.* AAPS Press, New York, 2004.

Communicating Pharmacometric Analysis Outcome

ENE I. ETTE and LEONARD C. ONYIAH

37.1 INTRODUCTION

Communication, in general, can be defined as information that enters a process and eventually leaves its inverse process (1). For instance, information is transmitted by speaking and received after processing by its inverse, hearing. *Communication* occurs if, and only if, information moves from the input to one process to the output from a second process, the latter process being the inverse of the former process. The information at the output of this inverse, receiving, process is known as *a communication* (1). Thus, communication involves the encoding and decoding of information.

Technical communication, in particular, is the process of gathering technical information and presenting it to a target audience in a clear, useful, accurate, comprehensive, grammatically correct, and easily understandable form (2). The term "technical" includes scientific, mechanical, chemical, legal, economic, medical, procedural, or other specialized information (2). Technical communicators study their audience and determine the best way to present the information. Should it be a table or a graph? The technical communicator reshapes this information so that the correct audience can access, understand, and use it. It stands to reason, therefore, that pharmacometric communication falls into the category of technical communication. Pharmacometric communication, therefore, is the encoding and decoding of pharmacometric knowledge.

It is important to effectively communicate the knowledge gained from the pharmacometric analysis of a clinical trial data outcome to the intended audience. Communication in this case would be most effective when answers to questions in the three categories: who, what, and why are addressed. The "who" category addresses the target audience to whom the communication is addressed. The "what" category deals with the subject of the communication, the content, and the style with which the message is conveyed or presented so that the target audience can

understand it. The "why" category deals with the reason for the communication, intended objective, and what is to be accomplished with the communication. Thus, it is important in communicating the knowledge extracted or created from a drug development data set to ensure that the message is appropriately addressed to the target audience, and the content and style of communication is appropriate for the audience, so that the intended objective of the communication is achieved. Effective pharmacometric communication can be achieved through words, tables, or graphics, which has been described respectively as infantry, artillery, and cavalry of the pharmacokinetic/pharmacodynamic defense force (3). These methods should be used to supplement one another, although the effectiveness of each depends on the contents of the message.

The subsequent sections deal with graphics in communication—incorporating graphical information processing, graphical perception, the effectiveness of graphical displays and numerical tables, the framework for graphical display, and graphical excellence and integrity. The emphasis of this chapter is the communication of pharmacometric knowledge and not different types of graphical displays (3) (see also Chapter 7 of this book).

37.2   GRAPHICS IN PHARMACOMETRICS COMMUNICATION

Graphs are analogous to written language: they communicate quantitative and categorical information, among others. Written language communicates thoughts, ideas, observations, emotions, theories, hypotheses, numbers, and so on. Graphical language is used extensively to convey information because it does so effectively. Quantitative patterns and relationships in data are readily revealed by graphs because of the enormous power of the eye–brain system to perceive geometrical patterns (4).

The power of a graph is its ability to enable one to take in the quantitative information, organize it, and see patterns and structure not readily revealed by other means of studying and presenting data (3, 5). The use of graphics in pharmacometrics communication should take into account how information in graphs is decoded, and special consideration should be given to the process of decoding the information in the graphs by the intended audience.

37.2.1   Graphical Information Processing

Process models for graphical perceptual processing have been proposed by cognitive psychologists (6–8), but statisticians and most pharmacometricians, on the other hand, like to think of the meaning of a graph as predefined. That is, if a graph is constructed properly its meaning will be self-evident. A simple summary of the distribution of an outcome variable displayed in a box plot may not be interpreted correctly by the intended audience. Viewers may read meaning into the width of the box even when it is constant in every box they see.

All stages in the graphical communication event are covered by a process model of graphical information processing, from the pharmacometrician who has knowledge/information to communicate, to the viewer who makes a judgment about that knowledge/information. The stages as described by Wilkinson (9) are:

- Quantitative/qualitative information
- Retinal image
- Decomposition in the visual cortex
- Integration and transformation via temporary storage in short-term memory and schemas accessed in long-term memory

37.2.1.1 Quantitative/Qualitative Information

Quantitative and categorical information may be part, but not all, of what the pharmacometrician wants to communicate to the viewer. The actual information in the data region may be a more complex arrangement of texture, form, edge, and other features. This distinction between the information the pharmacometrician desires to communicate and the actual organization of the graph is important because most of the formal arguments about graphs involve this stage. The pharmacometrician must select data features to highlight before constructing the graph. A single graph may not always reveal everything about the data. A well designed graph from a formal point of view may nonetheless be misperceived due to cognitive processes in the later stages. A review of how these come into play is discussed subsequently.

37.2.1.2 Retinal Image

The retina registers the initial image of the perceived graph. Because of lighting, viewing position, and other factors that can create different retinal images of the same graph, this image may differ in important ways from the physical image. Since black and white graphs with high contrast produce more constant retinal images under different lighting and viewing conditions, they are often preferable to ornate colored ones.

37.2.1.3 Decomposition in the Visual Cortex

By decomposing retinal images into features such as orientation and texture, these images are transformed in the visual pathway. The operations at this level are highly parallel and organized to extract spatial frequency, orientation, and other features needed to construct complex visual scenes.

Thus, it is important in graphical displays that high visual contrasts should be distinct entities on graphs. It has been surmised that maintaining the identity of separate graphic elements is often one of the truly challenging problems in making graphs, and it is an area of frequent failure. Exact overlap of plotting symbols prevents visually distinguishing distinct graphic elements. This is an extreme form of a more general problem. Even when all plotting points are distinct, symbols can partially overlap, and if too many points are crowded in a region of a plot, they can lose their identity. Whenever this becomes a problem, the use of unfilled circles or octagons as plotting symbols is recommended. These symbols can overlap and still maintain their identity. Squares and triangles do not share this property (3, 10).

37.2.1.4 Integration and Transformation Through Schemas

"Making a judgment about a graph involves integrating the features detected in the visual cortex by making use of a short term memory store—sometimes verbal, sometimes iconic—and schemas residing in long term memory. While the visual cortical operations are highly parallel, the operations at this stage are both parallel

and serial" (9). Temporary (less than half a minute) storage of information in order to perform serial operations is enabled by short-term memory. The viewer temporarily stores perceived information in order to construct higher-order comparisons such as scale references. Only a few (five to ten) distinct pieces of information can be stored simultaneously because of limitations in this store (9). Thus, higher-order interaction plots are difficult, if not impossible, to interpret because the number of comparisons required for understanding interactions increases exponentially. Thus, three-dimensional (3D) plots are difficult to interpret. Every effort should be made to reduce higher-order plots to two-dimensional (2D) plots. Contour, bubble, or *coplots* can be used in place of 3D or other complex plots. Appropriate 2D plots would enable the pharmacometrician to communicate effectively.

37.2.2   Graphical Perception

Numerical data can be displayed in different formats, but only some are well suited to the information processing capacity of human vision. The phrase *graphical perception* was coined by Cleveland and McGill (11) to refer to the role of visual perception in analyzing graphs. These authors studied several elementary visual tasks (such as discrimination of slopes, lengths of lines, or judging volume) relevant to graphical perception. They attribute the great advantage of graphical displays (e.g., scatterplots) over numerical tables to the capacity of the human vision to process pattern information globally at a glance. When there is compatibility between the task and the display type, perception of the judged characteristic is direct, requiring simpler or fewer mental operations. The merits of a graph, therefore, are task dependent and this should be borne in mind when choosing graphs for the communication of the outcome of a pharmacometric analysis.

37.2.2.1   *Effectiveness of Graphical Displays and Numerical Tables*
Tufte (12) makes the point that a graph can be more precise and revealing than a numerical display, and a graph can capture a large amount of information in a very small space. In comparing the effectiveness of data display using graphs and numerical tables, Legge et al. (13) found perceptual efficiencies to be very high for scatterplots, ≥60%. Efficiencies were much lower for numerical tables, 10%. Efficiency in the study referred to the performance of a real observer relative to that of an ideal observer. The ideal observer makes an optimal use of all available information (13). Performance with scatterplots was reported to have the hallmarks of a parallel process: weak dependence on viewing time. Processing of tables of numbers were found to be performed in a much more serial fashion. Their efficiencies dropped roughly with increasing information content in the tables and increased in rough proportion to viewing time. They concluded that entries in tables are processed sequentially at a fixed rate. Given enough viewing time, efficiency of information processing from tables could approach that of graphics (13). Thus, information content of tables should be kept to a minimum to allow efficient extraction of such information by the reader, or the audience in the case of oral presentation of the outcome of a pharmacometric analysis.

37.2.2.2   *Framework for Graphical Display*
As discussed in Section 37.2.2, the merits of a graphical display depend on the information chosen for display and the amount of effort that will be expended by

the reader in deciphering what is encoded in the graph (11, 14). In summarizing how sample size affects the power of a study, an integrated display such as a line graph would be a superior method of display. This is because in decoding the graph the reader would have to compare the power of the study for the different sample sizes and integrate that information to form his/her opinion. Judging change requires comparing quantities and integrating that information (11, 14). A line graph is more effective in conveying change than other types of display (11, 14). This is because the eye is focused on the physical slope of the line. Bar plots are also effective in conveying change (trends) in that the eye, in decoding change (or a trend) with bar plots, is tracing a perceived slope (11, 14). The effectiveness of a graph, therefore, depends on the amount of work that is to be performed by the reader in decoding the information contained in the graphical display.

In summarizing the results of a population pharmacokinetic study in which the effect of sample size on the bias and precision with which population pharmacokinetic parameters were estimated, Ette et al. (15) used line plots. A similar line plot display (Figure 37.1) of the effect of sample size and intersubject variability on the estimation of population pharmacokinetic parameters was created from an aspect of data generated in a simulation study performed to determine the performance of mixed designs in population pharmacokinetic studies (16). The plot shows the influence of intersubject variability on parameter estimation as sample size was varied. In the study, the effect of three different levels of intersubject variability, ranging from 30% to 60% coefficient of variation, and different sampling designs on the sample size required for efficient population pharmacokinetic parameter estimation were investigated. However, in Figure 37.1, data for only one of the designs are plotted to illustrate the effectiveness of a line plot.

FIGURE 37.1   A line plot of the effect of sample size and intersubject variability on the precision (expressed as percent mean absolute error—%MAE) with which central volume of distribution (V1) was estimated in a simulated population pharmacokinetic study in which a balanced sampling design was used.

The *coplot* (3, 17) is a powerful tool for studying how a response depends on two or more factors. It presents conditional dependence in a visually powerful manner. Two variables are plotted against each other in a series of overlapping ranges. This enables one to see how a relationship between two variables (y and x) change as a third variable (z) changes, that is, $y \sim x \mid z$. Thus, y is plotted against x for a series of conditioning intervals determined by z. Coplots may have two simultaneous conditioning variables, that is, $y \sim x \mid z1 \cdot z2$. Fadiran et al. (16) present a good example of the use of coplot in communicating the result of a population pharmacokinetics clinical simulation study. The presentation of the results from that study using coplots is a good example of using multipanel display to summarize the results of a simulated study. The use of the coplot allowed information from four variables to be communicated effectively using a 2D graphical display (see also Ref. 4).

In presenting data in a graphical display that requires attention to be focused on one variable, performance is better served by the use of more separated displays. The histogram and the box plot are examples of separated displays. The histogram (Figure 37.2) as used by Hale (18) in presenting the results of simulated randomized concentration-controlled trial with mycophenolate mofetil is a good example of the use of separated displays to convey information on a simulated study outcome. This plot compares simulation predicted trial outcomes and the actual trial result. The bars represent complete simulated trials using a developed simulation model. Outcomes to the right of the cutoff line are statistically significant, and the actual study outcome is shown. The actual trial value fell between the 80th and 90th percentile of the simulated results, which means that the actual trial outcome is

FIGURE 37.2   Frequency distribution of the test statistic for the primary analysis resulting from 500 completed simulated randomized concentration-controlled trials (RCCTs) with mycophenolate mofetil under "worst case" trial conditions, completed before real study initiation. The actual study outcome is shown, falling in the central portion of the distribution in the interval centered at 18.3 (From Ref. 18.)

FIGURE 37.3   A box plot of the distribution of area under the plasma concentration–time curve (AUC) for subjects who either failed or responded to antibiotic therapy in a clinical trial.

not unusual based on the simulation model, which reflects that the simulation model was a reasonable description of the trial process.

The box plot has proved to be a popular graphical method for displaying and summarizing univariate data, to compare parallel batches of data, and to supplement more complex displays with univariate information. Its appeal is due to the simplicity of the graphical construction (based on quartiles) and the many features that it displays (location, spread, skewness, and potential outliers). Box plots are useful for summarizing distributions of treatment outcomes. A good example would be the comparison of the distribution of response to treatment at different dose levels or exposure (as measured by area under the plasma concentration–time curve) as in Figure 37.3.

37.2.3   Graphical Excellence

According to Tufte (12), graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space. Five principles produce substantial changes in graphical design: above all else show the data. Maximize the data/ink ratio (i.e., the percentage of ink that shows data). Erase nondata ink. Erase redundant data ink. Thus, an excellent statistical graph consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays, therefore, should do the following (12):

- The graph should show the data above all else.
- The graph should persuade the viewer to think about the substance rather than about methodology, design, or something else.
- The graph must be a truthful representation of the data.
- The graph should encourage the eye to compare different pieces of data.
- The presentation of a large amount of data (information) should be made coherent to the reader or viewer.

- The revelation of levels of details of the data should be from a broad overview to the fine structure.

37.2.4 Graphical Integrity

It is important to ensure that the graphics used in the presentation of the results of the pharmacometric analysis of clinical trial data do not lie or mislead. The way to do this is to avoid visual distortion. That is, the pharmacometrician has to ensure that the visual representation of the data is consistent with the numerical representation. This is in terms of volume, area, and so on.

Human perception as well as actual changes must be taken into account. For example, the area of a circle is perceived by people to be decreasing less rapidly than the actual area: that is, the perceived area = (actual area)x with $x = 0.8 \pm 0.3$. Perceptions differ among people. Perception changes with experience, and perception is context sensitive. According to Tufte (12), two important principles should be borne in mind:

- The representation of the data should be proportional to the magnitude of the data.
- Clear and detailed labeling (annotation) should be used. The data should be explained on the graph and important data points should be labeled.

Tufte (12) defines a "lie factor" for graphics as

"lie factor" = size of the effect in the data/size of the effect shown on the graph

Any value for the lie factor other than one (practically within [0.95, 1.05]) indicates a distortion. This is a particular problem in using visual area, or even volume, to represent one-dimensional data. To ensure graphical integrity, the number of variable dimensions depicted should equal the number of dimensions in the data. Another means through which distortion can occur is through design and data variation. Observers assume that a scale will be regular, so variations can be used to distort the data. Changes in the design can cause observers to confuse this change with actual data change. Thus, data variation should be presented rather than variation in design. Context is important. Data should not be presented out of context. Leaving out information can create graphical distortion (12).

37.3 INFORMATION/KNOWLEDGE INTEGRATION

Since communication is defined as information that enters a process and eventually leaves its inverse process (1) and pharmacometrics communication is the encoding and decoding of pharmacometric knowledge, it stands to reason that attention should be paid to the encoding and decoding of a pharmacometric message. The message should be conveyed in an unambiguous manner so that information that enters the process leaves as its inverse. The audience should receive the exact message intended by the pharmacometrician. The pharmacometrician should not leave his/her audience to guess or come up with myriad interpretations that are

different from what was intended. This is why the previous sections emphasized the use of the right type of media to convey a pharmacometric message. It is important for the pharmacometrician to not leave the interpretation of his/her intended message at the mercy of the audience. Most often, the pharmacometrician is communicating the findings of his/her analysis to a drug development team that consists mostly of nonpharmacometricians. Succinctly encoding the pharmacometric communication for accurate decoding by the receiver cannot be overemphasized. Thus, pharmacometric communication should be done in clear, useful, accurate, comprehensive, grammatically correct, and easily understandable fashion.

37.4 SUMMARY

The importance of effectively communicating the pharmacometric knowledge extracted or created from a drug development data set to ensure that the message is appropriately addressed to the target audience is discussed. Pharmacometric communication is the encoding and decoding of pharmacometric knowledge, and effective pharmacometric communication can be effected through words, tables, or graphics. These methods should be used to supplement one another, although the effectiveness of each depends on the contents of the message. Excellent graphical display is recommended as the preferred method for communicating knowledge or information on a large amount of data. A graph can be more precise and revealing than a numerical display, and a graph can capture a large amount of information in a very small space. In communicating with graphics, it is important to bear in mind how the viewer or reader decodes the information. Thus, *graphical perception,* the role of visual perception in analyzing graphs, is an important component in communicating with graphics, and also in pharmacometrics communication. Graphical displays have a great advantage over numerical tables because of the capacity of human vision to process pattern information globally at a glance. It is recommended that tables should only be used when necessary, and information content of tables should be kept to a minimum to allow efficient extraction of such information by the reader. Pharmacometric communication should be done in clear, useful, accurate, comprehensive, grammatically correct, and easily understandable fashion. Failure to effectively communicate the outcome of a pharmacometric analysis puts at risk all of the data analysis efforts.

REFERENCES

1. R. M. Losee, Communication defined as complementary informative process. *J Info Com Library Sci* **5**(3):1–15 (1999).
2. C. McKitterick, *Technical Communication Defined*. `www.ku.edu/~techcomm/TC-Defined.rtf`. Accessed September 15, 2005.
3. E. I. Ette, Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial. *Ann Pharmacother* **32**:818–828 (1998).
4. E. I. Ette, C. J. Godfrey, S. Ogenstad, and P. J. Williams, Analysis of simulated trials, in *Clinical Trial Simulations*, H. C. Kimko and S. Duffil (Eds.). Marcel Dekker, New York, 2002, pp. 131–151.

5. E. I. Ette, P. Williams, H. Sun, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–54 (2001).

6. S. Pinker, A theory of graph comprehension, in *Artificial Intelligence and the Future of Testing*, R. Friedle (Ed.). Ablex, Norwood, NJ, 1990, pp. 73–126.

7. D. Simkin and R. Hastie, An information-processing analysis of graph perception. *J Am Stat Assoc* **82**:454–465 (1987).

8. S. M. Kosslyn, Understanding charts and graphs. *Appl Cognitive Psychol* **3**:185–226 (1989).

9. L. Wilkinson, Comment on W. S. Cleveland, A model for studying display methods of statistical graphics. *J Comp Graphical Stat* **2**:366–360 (1993).

10. W. S. Cleveland, A model for studying display methods of statistical graphics. *J Comp Graphical Stat* **3**:323–343 (1993).

11. W. S. Cleveland and R. McGill, Graphical perception and graphical methods of analyzing scientific data. *Science* **229**:828–833 (1985).

12. E. Tufte, *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2001.

13. G. E. Legge, Y. Gu, and A. Luebker, Efficiency of graphical perception. *Percep Psychophys* **46**:365–374 (1989).

14. J. S. Hollands and I. Spence, Judgement of change and proportion in graphical perception. *Human Factors* **34**:313–334 (1992).

15. E. I. Ette, H. Sun, and T. M. Ludden, Balanced designs in longitudinal population pharmacokinetic studies. *J Clin Pharmacol* **38**:417–423 (1998).

16. E. O. Fadiran, C. D. Jones, and E. I. Ette, Designing population pharmacokinetic studies: performance of mixed designs. *Eur J Drug Metab Pharmacokinet* **25**:231–239 (2000).

17. W. S. Cleveland, *Visualizing Data*. Hobart Press, Summit, NJ, 1993.

18. M. D. Hale, Using population pharmacokinetics for planning a randomized concentration–controlled trial with a binary response, in *The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose*, L. Aarons, L. P. Balant, M. Danhof, M. Gex-Fabry, U. A. Gundert-Remy, M. O. Karlsson, F. Mentre, P.L. Morselli, F. Rombout, M. Rowland, J. L. Steimer, and S. Vozeh (Eds.). Commission of the European Communities, European Cooperation in the Field of Scientific and Technical Research, Brussels, 1997, pp. l228–1235.

SPECIFIC APPLICATION EXAMPLES

Pharmacometrics Applications in Population Exposure–Response Data for New Drug Development and Evaluation*

HE SUN and EMMANUEL O. FADIRAN

38.1 INTRODUCTION

The drug development and approval process has presented increasing challenges to the pharmaceutical industry as well as to regulatory authorities in recent years. The United States Food and Drug Administration (FDA) recently issued a major report identifying both the problems and potential solutions to the daunting task of ensuring that the unprecedented breakthroughs in medical science are demonstrated to be safe and effective for patients as quickly and inexpensively as possible. Entitled *Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*, the report carefully examines the "critical path" of medical product development—the crucial steps that determine whether and how quickly a medical discovery becomes a reliable medical treatment for patients (1).

The report notes that despite notable advances in innovative fields of biomedical research as genomics, proteomics, and nanotechnology, there has been a downward trend in recent years in the number of innovative medical product applications to the FDA and its counterpart agencies throughout the world. While the number of new product applications and approvals was modestly higher in 2003, the fact remains that most of these new scientific fields are not yet having a fundamental impact on patient care. Although these and other problems are attributable to a variety of factors, the FDA's report focuses on one important cause—that new science is not being adequately harnessed to guide the drug development process in the same way that it is accelerating the discovery process. Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated, in a faster time frame, with more certainty, and at lower costs. In many

*The opinions expressed in this chapter are those of the authors and not those of the United States Food and Drug Administration (FDA).

cases, drug developers have been using the tools and concepts of the last century to assess this century's candidates, which results in the failure of the vast majority of investigational products that enter clinical trials. Often, product development programs are abandoned after extensive investment of time and resources and the path to market for successful candidates is long, costly, and inefficient, due in large part to the current reliance on cumbersome assessment methods.

It appears that pharmaceutical firms recognize the benefits of early attrition of potentially unsuccessful compounds but they also face a constant challenge of economic factors that influence the efficiency of drug development. Regulatory authorities, on the other hand, have been confronted with continued acceleration of the review process with the public expectation of uncompromised quality of safety and efficacy assessments. The Prescription Drug User Fee Act (PDUFA) (2) of 1992 and the Food and Drug Administration Modernization Act (FDAMA) (3) have provided the impetus and the much needed resources to implement substantial changes in the review process. As a result, review times have decreased significantly over the past 10 years (4).

In addition, there is an increased awareness within the FDA of the issues related to individualized drug therapy and the utility of possible confirmatory data from clinical trial sources other than the primary clinical safety and efficacy trials (FDAMA Section 115) (3). Amidst the "beat placebo" framework that has historically led to approval of doses that could be too high for individual patients, and the resultant safety issues, optimization of drug dosing has evolved into a primary goal of the drug development and review processes (5–8). This is because pharmaceutical sponsors have typically used the maximum safe dose (MSD) strategy rather than the minimum dose for satisfactory effect (MDSE) concept to identify doses for their Phase 3 program (5). This scenario has resulted in the present situation whereby lower effective doses have been used in clinical practice for several approved medications (9). Moreover, postapproval changes have been recommended for several approved medications in the United States (10) and Europe (11). In order to address this development, the FDA and the International Conference on Harmonisation (ICH) have published guidances for industry that provide an understanding of exposure response (E-R) and promote its application during drug development and regulatory review (12).

38.1.1 Understanding E-R Relationships

The drug approval process relies on evidence from the average drug performance in terms of safety and efficacy obtained from adequate and well controlled clinical trials in a select patient population. The approved doses or apparent exposures to drugs may be well beyond that needed for optimal benefit/risk and, in some susceptible individuals, these doses may even cause undesirable and sometimes serious side effects. The significance of careful characterization of the E-R relationship in drug development has long been realized by the FDA (13). An increased effort to better understand the relationship between drug exposure (as measured by dose, drug concentrations, or any other appropriate pharmacokinetic (PK) parameter) and the corresponding pharmacological response (either beneficial or adverse) has been made by the clinical pharmacology and clinical groups within the FDA (12).

E-R information can support the primary evidence of safety and/or efficacy. In some cases, E-R information can provide important insights that can allow a better understanding of the clinical trial data such as in explaining a marginal result on the basis of knowledge of systemic concentration–response relationships and achieved concentrations. Ideally, in such cases the explanation would be further tested, but in some cases this information could support approval. Even when the clinical efficacy data are convincing, there may be a safety concern that E-R data can resolve (12).

38.1.2   Applications of E-R Relationships

E-R information can sometimes be used to support use, without further clinical data, of a drug in a new target population by showing similar (or altered in a defined way) concentration–response relationships for a well understood (i.e., the shape of the exposure–response curve is known), short-term clinical or pharmacodynamic (PD) endpoint. Similarly, this information can sometimes support the safety and effectiveness of alterations in dose or dosing interval or changes in dosage form or formulation with defined PK effects by allowing assessment of the consequences of the changes in concentration caused by these alterations. In some cases, if there is a change in the mix of parent and active metabolites from one population to another (e.g., pediatric vs. adult), dosage form (e.g., because of changes in drug input rate), or route of administration, additional exposure–response data with short-term endpoints can support use in the new population, the new product, or new route without further clinical trials.

As noted in the ICH E4 guidance for industry entitled *Dose–Response Information to Support Drug Registration* (12), dose–response information can help identify an appropriate starting dose and determine the best way (how often and by how much) to adjust dosage for a particular patient. If the time course of response and the exposure–response relationship over time is also assessed, time-related effects on drug action (e.g., induction, tolerance, and chronopharmacologic effects) can be detected. In addition, testing for concentration–response relationships within a single dosing interval for favorable and adverse events can guide the choice of dosing interval and dose and suggest benefits of controlled-release dosage forms. The information on the effects of dose, concentration, and response can be used to optimize trial design and product labeling.

The various E-R study designs and their strengths and limitations have been extensively discussed in the ICH and FDA guidances (12) (see also Chapter 31 of this book).

The following examples are intended to illustrate the role that E-R plays in determining optimal dosing in various circumstances. The drug names have been masked to protect proprietary information; however, the principles of applying E-R information are clearly outlined.

38.2   DRUG ENHIBITOR

38.2.1   Objective

The objective of this analysis was to use the E-R relationship to find the optimal dose for Drug Enhibitor.

38.2.2   Background

Drug Enhibitor is an enzyme inhibitor being developed as an oral treatment for a functional disorder. Both 30 and 60 mg dosage strengths are available. The initial proposed dosing regimen was 60 mg once a day.

38.2.3   E-R Approach for Dose Selection

Several adequate and well controlled clinical efficacy and safety studies were conducted. There were three Phase 3 clinical studies, where 30 and 60 mg doses were compared with placebo. These studies were designed as randomized, double-blind, placebo-controlled, and parallel dose studies. The efficacy of the drug was measured as change from the baseline and all three efficacy parameters (E1, E2, and E3) were considered as primary evidence of efficacy.

The efficacy and safety data from Phase 3 studies and Phase 2 dose-ranging studies were analyzed to explore the E-R (efficacy and adverse events) relationship. Based on the E-R data, a dose with optimal efficacy and safety was recommended.

38.2.4   Population PK/PD Analysis

Population analyses using nonlinear mixed effects modeling were carried out with the data collected from three Phase 2 studies; the range of doses was 7.5 to 300 mg. Patient responses to a validated questionnaire developed for the functional disorder under investigation were used as endpoints in the PK/PD analyses. Response versus dose was modeled with the E_{max} model, which fit the data best. The parameter estimates from this model were used to calculate the probabilities of getting a certain clinical score. Based on the data illustrated in Figure 38.1, the probability of achieving an acceptable efficacy measure (defined as >15% improvement in score) increased with dose. The results demonstrated that the probability of reaching the highest score for the PD endpoint was not significantly higher at 60 mg when compared with the 30 mg dose.

Overall, it appeared that the population response reached a plateau at approximately 30 mg; therefore, administration of doses greater than 30 mg did not provide additional benefit. This conclusion was consistent with the findings resulting from the efficacy data reported in the Phase 3 trials.

Based on the pooled data from the three Phase 3 studies, there was a trend toward dose-dependent increases in the frequency of drug-related adverse events, particularly dyspepsia, myalgia, and back pain, although the overall absolute incidence was between 0.5% and 10%. This was dependent on the event and drug dose.

Given the results of the pharmacometric analysis, a starting dose of 30 mg was recommended for patients with the functional disorder who are otherwise healthy.

With regard to special populations, systemic Enhibitor exposure was approximately twofold higher for the parent drug and three- to fourfold higher for the major metabolite in subjects with mild and moderate renal impairment following administration of the 30 mg dose. The increased systemic Enhibitor exposure in this population was associated with significant increase in the incidence of drug-related

FIGURE 38.1 Estimated probability of achieving a given score of efficacy measure for Enhibitor. In this plot, the efficacy measure is the probability of achieving a desired clinical efficacy score (or percentage of patients who achieve the desired efficacy score). Two representative curves are displayed—one for when the desired efficacy score is 5 (solid line), and the other when the desired efficacy score is 3 (dashed line). When the desired clinical efficacy score is 3 (i.e., clinically satisfactory), the probability of clinically achieving this score plateaued at 20 mg of dose (the probability is 93%). Increasing dose will not further increase this probability. If the desired efficacy score is 5 (clinically very satisfactory), the probability of achieving this score plateaued somewhere beyond the 75 mg dose, but the added benefit from 30 mg to 60 mg is relatively small (increased from 60% to 66%). Therefore, doses greater than 30 mg would not likely add significant clinical benefit.

adverse events. Due to the increased incidence of adverse events in moderate renal impairment subjects, the exposure of Enhibitor was not investigated in patients with severe renal impairment. It should be noted that all the pivotal Phase 3 trials for Enhibitor excluded patients with clinically significant renal failure.

38.2.5 Results

Based on the E-R data analysis, there were no significant differences ($p > 0.05$) in response between 30 and 60 mg doses. Both 30 and 60 mg doses were statistically significantly better than placebo. There did not seem to be any additional benefit with the 60 mg dose compared to the 30 mg dose.

Based on the increased Enhibitor exposure of parent drug and the metabolite in patients with moderate renal impairment, and the associated increase in incidence of adverse events, a dose of 15 mg was recommended in this patient population.

38.2.6   Conclusions

The sponsor had requested approval for only a 60 mg dose. Based on the E-R analysis of the data from Phase 2 and Phase 3 studies, a lower starting dose of 30 mg in patients with the functional disorder who are otherwise healthy was recommended. Based on the evidence of increased Enhibitor exposure and adverse events in renal impairment patients, it was concluded that dosing in patients with moderate renal impairment should not exceed 15 mg and the drug should be contraindicated in patients with severe renal impairment.

38.3   DRUG BOTANI

38.3.1   Objective

The objective of this analysis was to determine the lowest effective dose of a botanical drug product, Botani, indicated for a chronic disorder. The drug was to be given chronically over many years. The sponsor requested approval of a dose that is similar or better than a marketed comparator, Compara.

38.3.2   Data Resource

Botani and Compara were both administered once daily for 8 weeks. There was a run-in period and follow-up phase. Data from Phase 2 dose–response studies were available. The E-R relationship was characterized in two Phase 2 clinical trials with the daily doses of 0, 3, 7.5, 15, 30, 60, 120, and 240 mg. Percent change of plasma R from baseline ($\Delta R\%$) was used as a PD clinical endpoint.

38.3.3   Data Transfer

Data was transferred from SAS transport file to ASCII format using StatTransfer. Final data sets for the consequent analysis were saved in S-Plus data structure.

38.3.4   Data Analysis and Results

Data analyses included data visualization, nonparametric statistical analysis on observations (data from study 1 only), and parametric analysis with nonlinear mixed effects modeling.

38.3.4.1   Data Visualization
Various plots were generated to check $\Delta R\%$ versus treatment time relationship, $\Delta R\%$ versus dose relationship, variability, relative potency between Botani and Compara, and the effect of covariates on the response variable. Botani exposure–response plots indicated that plasma R gradually drops and essentially reaches plateau over the 6 week treatment time. The profiles of $\Delta R\%$ over time showed similar pattern for both Botani and Compara (Figure 38.2). After approximately 4 weeks of treatment time, percent change of endpoint ($\Delta R\%$) reaches 86–90% of the maximum effect for a given dose. This was observed for all dose levels studied (3–240 mg). It did not appear that any covariates (age, sex, baseline R, etc.) had any

FIGURE 38.2 Plots of $\Delta R\%$ over time by dose and treatment for study 1 (a comparative study of Botani and Compara). Data show treatment effect over time and visual comparisons on relative potency of Botani and Compara. At all doses, the $\Delta R\%$ gradually increases and reaches steady states. When the drug is withdrawn, R slowly returns to the baseline level. The new drug, Botani, produced more $R\%$ reduction than Compara as shown at the 30 mg dose. In addition, there is a clear dose–response relationship for Botani.

effect on $\Delta R\%$. Botani produced up to approximately $4\times R$ lowering effect when compared with Compara, in both studies (Figure 38.3). As shown in Figure 38.3, the maximum effect for Botani was approximated at the 30 mg dose.

38.3.4.2 Data Analyses

Botani's dose–response and concentration–response analyses were conducted. In the analyses, no effort was made on checking per protocol (PPT) data, but the focus was on the intention-to-treat (ITT) data. Analyses focused on the dose–response data in study 2 and study 1 data were used for external model validation (see Chapter 8 for a detailed discussion on model validation). Data from other studies were not modeled. Since concentration–effect data from study 2 did not appear to offer any advantage when compared with the dose–response data, it was examined without detailed modeling analyses. Statistical comparisons between $R\%$ reductions at a given Botani dose versus $4\times$ Compara dose resulted in the 90% CI of the mean ratios of the two treatments falling within 80–125%, indicating that the mean observations or parameters for the two drugs at doses $4\times$ different were equivalent.

Some statistical analysis results of Botani's E-R data are presented in Tables 38.1. The $\Delta R\%$ reduction at week 6 was not significantly different from that at week 4. Results of mean (and SD) of $R\%$ reduction of study 1 and study 2 were

FIGURE 38.3   Plot of ΔR% versus dose by treatment. Data show visual comparisons on relative potency of Botani and Compara at all dose levels. In all doses, the new drug Botani produces a greater ΔR% than Compara. The dose–response relationship is clearly demonstrated. The data points are observed ΔR%; the lines are links of the mean values at various doses.

TABLE 38.1   Mean ΔR% at Various Dose Levels for Botani at Week 6

Botani Dose (mg/day)	Mean ΔR% at Week 4	Mean ΔR% at Week 6	Study 1 Observed ΔR% at Week 6 (Meanand SD)	Study 2 Observed ΔR% at Week 6 (Meanand SD)	Approximate Minimum ΔR% at Week 6 in About 85% of Patients
0	−1.38	−1.99	−0.598 (7.18)	−1.318 (6.58)	—
3	−36.67	−38.32	−35.23 (8.81)	—	−26
7.5	−39.16	−40.93	−41.57 (9.27)	—	−28
15	−42.55	−44.47	−44.63 (7.16)	−41.6 (9.94)	−32
30	−47.39	−49.53	−49.43 (17.0)	−49.95 (10.67)	−39
60	−53.09	−55.9	−54.26 (11.9)	−52.21 (9.86)	−43

very consistent. Also, at the 15 mg dose, R% reduction is more than 30% in 85% of subjects, yielding a clinically significant therapeutic outcome.

38.3.4.3   Nonlinear Mixed Effects Modeling Analyses

The objective of this analysis was to integrate all of the above information for making a final recommendation on optimal therapeutic dose of Botani. Nonlinear mixed effects modeling analyses were conducted only on dose–response data from study 2 because of its completeness at multiple dose levels and larger number of subjects. A total of 374 subjects with 1816 observations were included in the data analyses. An inhibitory effect E_{max} model describes the response–time relationship (at a given

dose) and response–dose relationship (at a given time point) very well. However, concentration–response data did not offer any advantage over dose–response data in these modeling analyses. The nonlinear mixed effects analyses were conducted using Pharsight WinNonMix software. Model-building criteria for adding covariate effects were based on objective function change by more than 30 units ($p < 0.001$) and the examination of diagnostic plots. An Inhibitory Maximum Effect function available in WinNonMix was used for both time effect and dose effect on $\Delta R\%$. The FOCE BLOCK method was applied. Intersubject variability was modeled as lognormal distribution and residual error was modeled unweighted.

The final nonlinear mixed effects model was (applicable to the dose range 3–240 mg only)

$$\text{Max observed } \Delta R\% = \text{Baseline (day 0)} + \text{Treatment time effect} + \text{Dose effect} + \text{Drug effect}$$

$$\text{Baseline at day } 0 = -1.99\%$$
$$\Delta R\% \text{ over time} = \text{Baseline} - E_{\max} \cdot (1 - \exp(-K \cdot \text{Time})) \tag{38.1}$$

where $E_{\max}$ is a regression parameter to describe the maximum effect at time infinity and is drug and dose dependent:

$$K = 0.1657 - 0.00789 \cdot (\text{trt-1})$$
$$E_{\max} = -39.94 - 38.40 \cdot \text{Dose}/(16.43 + \text{Dose}) + 10.03 \cdot (\text{trt-1}) \tag{38.2}$$
$$\text{trt} = 1 \text{ if drug is Botani; trt=2 if drug is Compara}$$

The model excellently predicted individual $\Delta R\%$ over time as indicated in a set of goodness-of-fit diagnostic plots (not shown) for study 1. The predicted mean values have less than ±4% error. Post hoc individual prediction for all 374 subjects was excellent (see the appendix for details and individual predictions). The χ^2 statistics confirmed that there were significant treatment effects, dose effects, and time effects on $\Delta R\%$ profiles for Botani and Compara. Based on the model, it was concluded that: (a) Botani offered superior potency to Compara; for example, 30 mg Botani produced approximately equal degree of $R\%$ lowering effect as 120 mg Compara; (b) at the same dose level (e.g., 30 mg), Botani would produce an additional 20% R lowering effect than Compara (e.g., 49.95% drop in R for Botani versus 37.87% drop in R for Compara); and (c) at least 4 weeks of treatment time are needed to approximate the corresponding maximum effect for all doses.

38.3.4.4   Results
E-R analyses indicated the following. Significant reductions in plasma R were seen within 1 week of therapy and most of the total effect was achieved by 4 weeks. The extent of $R\%$ lowering was dose-related and was in the range of −35% to −60% for 30–60 mg doses. Although the sponsor had proposed to market 30, 60, 120, and 240 mg doses, the E-R data showed that doses lower than the 30 mg also reduced plasma R levels significantly.

A 7.5 mg dose reduced the plasma R concentrations by a mean of 40% and a 3 mg dose, the lowest dose studied, reduced plasma R by approximately a mean of 35%. In addition, prediction based on nonlinear mixed effects modeling

FIGURE 38.4   Simulated population distribution plots of $\Delta R\%$ for doses 3, 15, and 30 mg. Data are from the NONMEM model prediction. If $\Delta R\% = 20\%$ is assumed to be a clinical effective benefit marker, the increase in the number of patients who will benefit from the treatment is small when the dose was increased from 3 to 30 mg.

indicated that the approximate minimum $\Delta R\%$ at 6 weeks in about 85% of patients who took the 3 mg, 7.5 mg, 15 mg, and 30 mg doses were –25%, –30%, –35%, and –40%, respectively. The efficacy measure appeared to plateau at the 120 mg dose with little additional benefit achieved when titrated from 3 to 30 mg (Figure 38.4). This absence of additional benefit was coincident with greater risk for myopathy and rhabdomyolysis observed at higher doses.

38.3.5 Conclusions

Although the sponsor had proposed to market 30, 60, 120, and 240 mg doses, the E-R data indicated that doses lower than the 30 mg were efficacious. Considering the potential toxicity at higher doses and the increased risk of greater exposure in the patient subgroup (e.g., renally impaired), it was recommended that the optimal starting dose be lower than 30 mg (5 and 15 mg doses). The safe use of drug Botani would require the availability of low-dosage strengths (e.g., for patients treated with other interacting drugs, patients with renal impairment, or the elderly).

38.4 DRUG AMICID

This Investigational New Drug (IND) application for this drug was initially submitted in July 2001. Drug Amicid is a water-soluble hormone antagonist. The product is formulated as a lyophilized powder, reconstituted with sterile water and mannitol for subcutaneous and intramuscular injection, or with sterile 5% glucose solution for intravenous infusion. Drug Amicid is relatively safe. The single dose tested was up to 360 mg without adverse effects. There was a very low ADR rate.

The sponsor conducted several Phase 2 trials with various subcutaneous injection volumes, dose, dosing interval, and length of durations. With these data, the sponsor

conducted comprehensive modeling and simulation work to design the optimal dose and dose regimen for their planned Phase 3 trials, including monthly and trimonthly dose regimens. Sponsor requested an end of Phase 2a (EOP2A) meeting to receive the FDA's input on the modeling & stimulation (M&S) performed, and the Phase 3 clinical program for the drug.

Given the amount of trough concentration data available, the data were sub-divided into 1 ng/mL bins. The corresponding percentage of subjects within each concentration bin who met a clinical efficacy measure criterion was calculated. The clinical group desired the efficacy rate to be greater than 90%. Figure 38.5 is a composite plot of such data. It indicates that if subjects maintained Amicid trough concentrations within 5–6 ng/mL, the clinical success rate would be about 92%. Since controlling mean trough drug concentrations in a population was more practical, it was necessary to incorporate variability, factoring in the concentrations and, with the above criterion, the percentage of subjects who had Amicid trough concentrations lower than 5 ng/mL at various mean Amicid trough concentrations were generated and plotted (see Figure 38.6). From the figure it can be observed that when the mean concentration was 10 ng/mL, the percentage of subjects with trough concentrations less than 5 ng/mL was less than 1%, and the 10 ng/mL average concentration appears to be the inflection point on the curve. When the mean plasma drug concentration was less than 10 ng/mL, the percentage of subjects who

FIGURE 38.5   Percent success rate versus drug concentration. Data were generated by binning the concentrations (see text for explanation). Different symbols show data from different treatment times. Tolerance was not observed. If all subjects have trough Amicid concentration of 5 ng/mL, the success rate would be about 92%. Due to concentration variability in a population, a mean concentration in a population of 7.5 ng/mL corresponds to a success rate ranging from 70% to 97% (mean 94%, as shown in the darker square), and a success rate of 92–97% (mean 96.4%, the lighter square) for a corresponding mean concentration of 9.5 ng/mL.

FIGURE 38.6   The percentage of subjects who have trough Amicid concentrations lower than 5 ng/mL at various mean Amicid concentrations. When mean Amicid concentration is greater than 10 ng/mL, the percentage of subjects whose Amicid concentration is less than 5 ng/mL is about 2–4%. However, a further increase in the mean concentration will not offer much added benefit to reduce the percentage of subjects with trough Amicid concentrations less than 5 ng/mL.

would not be able to maintain Amicid trough concentrations greater than 5 ng/mL increased in an almost linear manner. Furthermore, an increase in mean concentration beyond 10 ng/mL would not offer any substantial improvement in decreasing the fraction of subjects with plasma Amicid concentration lower than 5 ng/mL.

Based on the above information, trial design was simulated using the sponsor's PK model with the criterion that the mean drug concentration remains on or slightly above 9.5 ng/mL. Furthermore, various dose–dosage regimen combinations were simulated to recommend the optimal dose–dosage regimen combination. The sponsor accepted the Agency's recommendation for the dose–dosage regime combination to study in their Phase 3 clinical trials.

38.5   DISCUSSION

According to the ICH-E4 guideline (12) and certain FDA guidances (12), a knowledge of the relationships among dose, drug concentration, and clinical response (desirable and undesirable effects) can (a) guide the selection of an appropriate starting dose, (b) guide the selection of an adjusted dose for patient subgroups, and (c) identify a dose beyond which increases would be unlikely to provide added benefit or would produce unacceptable side effects.

For a pharmaceutical product that is prescribed to a patient, the ideal goal is for the patient to attain maximum efficacy (i.e., maximum benefit) with minimal adverse events (i.e., limited risk). These examples illustrate how E-R relationships can guide the selection of not only an efficacious dose but an optimal dose as well. Using traditional approaches for assessing E-R relationships and population PK/PD methodologies, optimal doses for Drug Enhibitor and Drug Botani were chosen from several available efficacious doses. Without such information, higher than necessary doses may have been approved with associated greater risks of adverse events. In the case of drug Amicid, an understanding of the E-R relationship led to the recommendation of the right dose to be studied in the Phase 3 clinical program. Dose finding in Phase 3 is also important for dosage optimization (14).

38.6   SUMMARY

In conclusion, these examples taken from two recently submitted NDAs to the FDA and an EOP2A meeting between FDA and a sponsor demonstrate that E-R information can significantly contribute to a better understanding of optimal doses and dosage regimens. It is also important that dose finding not only occur in Phase 2 but that it continues into Phase 3 in order to optimize dosing.

REFERENCES

1. *Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products.* FDA, Rockville, MD, 2004. `http://www.fda.gov/oc/initiatives/critcalpath/whitepaper.html`.

2. PDUFA of 1992. `http:/www.fda.gov/oc/pdufa`.

3. FDAMA of 1997. `http://www.fda.gov/cder/fdama`.

4. CDER Annual Report. http://www.fda.gov/cder/reports/rtn/2001/rtn2001.htm.

5. J. L. Lesko, M. Rowland, C. C. Peck, and T. F. Blasche, Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *J Clin Pharmacol* **40**:803–814 (2000).

6. C. C. Peck, Drug development: improving the process. *Food Drug Law J* **52**:163–167 (1997).

7. P. Roland, The contribution of clinical pharmacology surrogates and models to drug development—a critical appraisal. *Br J Clin Pharmacol* **44**:219–225 (1997).

8. H. Derendorf, L. J. Lesko, P. Chaikin, W. A. Colburn, P. Lee, R. Miller, R. Powell, G. Rhodes, D. Stanski, and J. Venitz, Pharmacokinetic/pharmacodynamic modeling in drug research and development. *J Clin Pharmacol* **40**:1399–1418 (2000).

9. J. S. Cohen, Dose discrepancies between the PDR and the medical literature, and their possible role in the high incidence of dose-related adverse drug events. *Arch Intern Med* **161**:958–964 (2001).

10. J. Cross, H. Lee, A. Westelinc, J. Nelson, C. Grudzinska, and C. Peck, Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999. *Pharmacoepidemiol Drug Safety* **11**:439–446 (2002).

11. E. R. Heerdink, J. Urquhart, and H. G. Leufkens, Changes in prescribed drug doses after market introduction. *Pharmacoepidemiol Drug Safety* **11**:447–453 (2002).

12. http://www.fda.gov/cder/guidance/index.htm:

ICH E4: *Dose–Response Information to Support Drug Registration* (1994).

Exposure–Response Relationships: Study Design, Data Analysis and Regulatory Applications (2003).

Providing Clinical Evidence of Effectiveness for Human Drug and Biologic Products (1998).

ICH E9: *Guidance on Statistical Principles for Clinical Trials* (1998).

13. R. Temple, Dose–response and registration of new drugs, in *Dose–Response Relationships in Clinical Pharmacology*, L. Lasagna, S. Eri, and C. A. Naranjo (Eds.). Elsevier, Amsterdam, 1989, pp. 145–167.

14. L. B. Sheiner, Learning vs. confirming in clinical drug development. *Clin Pharmacol Ther* **61**:275–291 (1997).

APPENDIX 38.1

```
WINNONMIX NONLINEAR MIXED-EFFECTS ESTIMATION PROGRAM (V2.0.1)
                    Core Version 28NOV2000
Listing of Input Commands:


MODEL
DNAMES ID~1 OBDAY~2 PD~3 WEEK~4 TRT~5 DOSE~6 TRT2~7 DAY~8
METHOD 1 /R
MINIMIZATION 0 /STEP=1
DINCREMENT 0.001
NPOINTS 100
ITERATIONS 100
CONVERGENCE FUNC /TOL=0.0001
SUBJECT ID
XNAME DAY
YNAME RP
STDERR 0
MIXEFFECTS
  BASE=BASE1_0                        ;Baseline term in equ 38.1
  K=(K0_1+K0_2*(TRT2-1))*EXP(K0_ETA0)   ;K term in equ 38.1
  E=(E0_0+E0_1*DOSE/(E0_2+DOSE)+E0_3*(TRT2-1))*EXP(E0_ETA0);Emax
term in equ 38.2
END
INITIAL
  1: (-2) (0.19) (0.002) (-42) (-35) (15) (8)
END
NOBOUND
VFUNCTION IDENTITY
OMEGA K0_ETA0 E0_ETA0
END

Command Parsing Completed.
Data Input Finished.
```

```
                 Ordinary Least Square Estimates of Fixed Effects


       EMAX_0        EC50_0        EC50_1         E0_0          E0_1          E0_2
  -2.2878E+00  3.9733E+00  -4.1081E-01  -5.2616E+01  -2.2337E-01  9.3982E+00



                      Initial Estimates of Random Effects
                                (For All ID)


   EC50_ETA0       E0_ETA0
   0.0000E+00  0.0000E+00



                   Initial Estimates of Covariance Parameters


SIGMA^2:                                          5.6588E+01
Covariance Matrix of Random Effects:
             EC50_ETA0        E0_ETA0
EC50_ETA0  1.3998E+00
E0_ETA0                     7.6006E-02


Computation of Initial Estimates Completed.



                        REML Estimation Iteration History


Iteration            Objective        Criterion
1                  21711.9269            1.0000
    EMAX_0          EC50_0        EC50_1         E0_0          E0_1          E0_2
   -1.9804E+00  4.2012E+00  1.3577E-01  -5.1956E+01  -2.7583E-01  8.8429E+00


Iteration            Objective        Criterion
2                  21711.9269            0.0000
    EMAX_0          EC50_0        EC50_1         E0_0          E0_1          E0_2
   -1.9804E+00  4.2012E+00  1.3577E-01  -5.1956E+01  -2.7583E-01  8.8429E+00


Convergence Achieved.
Model Estimation Completed.



                             Class Level Information


Class   Levels Values


ID       374 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
             ......373 374
```

```
                      Model Fitting Information

Description                              Value


Number of Subjects                          374
Total Observations                         1816
Minimum Objective Function Value      21711.9269
REML Log Likelihood                   -6635.7629
Akaike's Information Criterion (AIC)  13287.5257
Schwarz's Bayesian Criterion (SBC)   13331.5344
-2 * REML Log Likelihood             13271.5257

                   Solution For Fixed Effects


Parameter      EMAX_0       EC50_0       EC50_1         E0_0         E0_1
Estimate  -1.9804E+00  4.2012E+00  1.3577E-01  -5.1956E+01  -2.7583E-01
StdError   4.3267E-01  2.8473E-01  4.6437E-01   1.1187E+00   2.1160E-02

Parameter         E0_2
Estimate   8.8429E+00
StdError   1.3235E+00

                  Covariance Parameter Estimates

Parameter           Estimate           StdError
SIGMA^2            6.1175E+01         6.3153E+00

Variance/Covariance Of Fixed Effects:

             EMAX_0         EC50_0        EC50_1           E0_0           E0_1
EMAX_0    1.8720E-01
EC50_0   -5.4909E-03     8.1072E-02
EC50_1   -3.4663E-02    -7.3546E-02    2.1564E-01
E0_0      1.1176E-02    -1.3797E-01    1.3743E-01    1.2515E+00
E0_1      1.2942E-04    -2.4125E-04   -2.1655E-04   -1.3464E-02    4.4774E-04
E0_2      2.6031E-02     1.3654E-01   -3.3834E-01   -6.9148E-01   -3.4679E-03

                E0_2
E0_2      1.7517E+00

Correlation Of Fixed Effects:
             EMAX_0         EC50_0        EC50_1           E0_0           E0_1
EMAX_0    1.0000E+00
EC50_0   -4.4571E-02     1.0000E+00
EC50_1   -1.7252E-01    -5.5624E-01    1.0000E+00
E0_0      2.3091E-02    -4.3316E-01    2.6456E-01    1.0000E+00
E0_1      1.4137E-02    -4.0042E-02   -2.2039E-02   -5.6878E-01    1.0000E+00
E0_2      4.5458E-02     3.6231E-01   -5.5050E-01   -4.6703E-01   -1.2383E-01


                E0_2
E0_2      1.0000E+00
```

```
Variance/Covariance Of Random Effects:
             EC50_ETA0        E0_ETA0
EC50_ETA0    2.6183E-01
E0_ETA0                       2.0195E-02


Standard Error of Variance/Covariance Of Random Effects:
             EC50_ETA0        E0_ETA0
EC50_ETA0    6.0387E-02
E0_ETA0                       2.9767E-03


Correlation Of Random Effects:
             EC50_ETA0        E0_ETA0
EC50_ETA0    1.0000E+00
E0_ETA0                       1.0000E+00


Variance/Covariance Of Individual Estimates:

            EMAX            EC50               E0
EMAX    0.0000E+00
EC50    0.0000E+00    7.7233E+01
E0      0.0000E+00    6.6950E+00    7.6264E+01


  Solution For Random Effects

ID      EC50_ETA0        E0_ETA0
1       -4.3592E-01    1.4015E-01
2       -1.0799E+00    2.9815E-01
37      -2.2513E-01   -2.8100E-02
38      -5.8753E-01    8.0395E-02


  Individual Parameter Estimates

ID          EMAX            EC50              E0
1       -1.9804E+00    2.8046E+00   -4.9600E+01
2       -1.9804E+00    1.4730E+00   -5.8090E+01
373     -1.9804E+00    3.7591E+00   -4.0335E+01
374     -1.9804E+00    5.1892E+00   -3.3586E+01


  Estimates Of Secondary Parameters

ID        EMAX-E0
1       4.7620E+01
2       5.6110E+01
373     1.2392E+00
374     1.0547E+00


Program Completed (total time used: 00:02:48.45).
Normal Ending.
```

Pharmacometrics in Pharmacotherapy and Drug Development: Pediatric Application

EDMUND V. CAPPARELLI and PAUL J. WILLIAMS

39.1   INTRODUCTION

The use of medication by children is widespread, with 20% of school-age children receiving one or more prescription drugs each year. Given the extensive use and altered pharmacokinetics resulting in a wide range of drug exposures, it is not surprising that recent history is replete with tragic consequences of drug administration to pediatric patients when pediatric clinical pharmacology information is lacking. To promote healthy children, it is of paramount importance that optimal dosing strategies be determined. The development of dosing strategies in this population is, in general, more complex than adults because of the diversity in the pediatric population, which is much greater than adults. For example, this patient population can range from several hundred grams to over a hundred kilograms and there is diversity of maturation where organs of elimination have varying degrees of functionality at different ages. Many biomarkers also have age-dependent ranges and therefore can exhibit significant within-subject maturational changes even over the course of a study. Size, age, ongoing growth, altered disease progression, and maturation are only several of the factors that impact dosing strategy in pediatric patients. Comprehensive and well defined development of pharmacokinetic (PK), pharmacodynamic (PD), and outcomes models are a necessity for the generation of optimal pediatric therapy.

By the mid-1960s the problem of administering inappropriate drug doses to pediatric patients had been clearly documented. In 1959 reports of the gray baby syndrome in neonates were published documenting the toxicity of chloramphenicol when adult doses were "miniaturized" to infants without consideration of maturation differences. The resulting deaths occurred because neonates have immature glucuronyl transferase activity necessary for the biotransformation of chloramphenicol and therefore accumulation occurred. When chloramphenicol

accumulated in the neonate's bloodstream, it caused hypotension, cyanosis, and often death.

Many of the laws defining the US Food and Drug Administration (FDA) activities have been developed as a result of therapeutic misadventures in infants and children. The 1938 Amendment to the Federal Food and Drug Act was enacted as a direct result of the distribution of a drug, sulfanilamide elixir, that killed 107 children due to the use of diethylene glycol as a solvent. Passage of the Kefauver–Harris FDA Amendment in 1962 was prompted by the tragic malformations seen in babies who were exposed to thalidomide in utero. Even with these additional safety regulations in place, infant deaths occurred from gasping syndrome due to benzyl alcohol use as a preservative and the limited ability of infants to metabolize and eliminate this "nontoxic" preservative (1). Beyond PK changes, the growth and development that occur during childhood can be adversely affected. Recently, it has been recognized that use of systemic steroids in infants with respiratory distress results in stunted head growth and cerebral palsy (2). The clinical pharmacology of drugs administered to children cannot be extrapolated from adult data on absorption, metabolism, and excretion alone. Although this concept is now generally accepted, the special characteristics of the pharmacology of drugs in children continues to be underappreciated.

Although the mantra for the National Institute for Childhood Diseases (NICHD) rightly proposes that children are not miniature adults, they are not Martians either. As the knowledge of pediatric pharmacology has increased, some extrapolations can be made with a reasonable level of certainty. With this in mind, the FDA has recently published the draft of *Guidance for Industry: General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products* (3). For ethical and logistic reasons, pharmacometric studies in children almost always occur after the characterization of a drug in adults and thus pediatric trials can utilize existing knowledge, including the use of preclinical data, in their design.

Distribution characteristics and elimination pathways show commonality across ages but their relative importance may differ. Therefore, a drug that is a substrate for a particular drug metabolizing enzyme isoform in adults will remain a substrate for that isoform in children. However, in extreme cases, pathway switches may occur such that the primary route of elimination in adults may be undeveloped in infants to such an extent that an alternative pathway predominates. Caffeine is an example of this, where it is primarily renally eliminated in infants due to the almost completely undeveloped CYP 1A2 biotransformation pathway. However, use of the growing knowledge of the developmental pattern of the physiologic processes that are important for a drug's disposition can be used to predict PK behavior in pediatric populations.

While there is a high degree of similarity across age groups relative to enzyme presence and activity, one important exception exists—the CYP 3A system. In adults, the CYP 3A4 isoform is responsible for metabolism of more drugs than any other enzyme but this isoform is essentially absent in infants with the primary isoform in this family being CYP 3A7, which is not found in adults. While there is overall homology in substrates for these two enzymes, there are differences that may lead to unexpected developmental changes in the pharmacokinetics for drugs that are CYP 3A substrates.

In pediatrics, the maturation changes in pharmacodynamics can be of a greater magnitude than PK differences, yet pediatric-specific PD models remain infrequently and poorly defined. Potential maturational PD differences can be pronounced depending on the physiologic or pathophysiologic system involved. This is especially true for systems that undergo extensive maturation after birth and for the least developed infants, those born premature. The immune and central nervous systems are particularly prone to altered pharmacodynamics due to their extensive postnatal development. GABA pathways, which produce inhibitory responses in adults, do not fully develop until 10 years of age and may be involved in excitatory pathways in preterm infants, which has been clinically manifested by clonic responses to lorazepam. In older infants paradoxical excitatory responses to nonsedating antihistamines may also be a manifestation of a similar phenomena. The immune system response to potential pathogens is grossly underdeveloped in infants and acceptable anti-infective PD targets in adults may not be adequate for infants. It also puts infants at greater risk for therapies whose primary adverse effects include immunosuppression. A further example of significant postnatal development is the autonomic nervous system that in the first year of life impacts the pharmacodynamics of drugs that affect the cardiovascular and gastrointestinal systems.

The exposure–response surface (see Chapters 8 and 32) is rarely mapped in children, thus limiting extrapolation of therapy from adults to children. Models linking biomarkers to patient outcomes—outcomes link models—are virtually absent in pediatric pharmacometrics. There is no reason to believe that PK/PD-biomarker-surrogate-outcomes linkage in children should uniformly parallel adults; therefore, this is an area where there is a great need for further model development. Population methods can play an important role for filling in these critical pieces to determine optimal pediatric therapy.

Developmental differences, disease presentation, disease progression, and comorbidities also need to be considered when determining pediatric pharmacotherapy. Even when the mechanism of action and PD response surface may be similar between pediatric and adult populations, differences in therapy may be indicated based on disease progression. For example, hypertension rarely presents as primary finding in children but most frequently as secondary to renal disease or other processes, which frequently impact the pharmacologic goals of therapy. HIV infection and AIDS will result in a 50% 2-year mortality in untreated infants yet typically takes 10 years in adults to wear down the immune system to the point at which opportunistic infections and AIDS take hold. Thus, therapeutic targets must account for these differences especially if these therapies will be used for chronic conditions.

The application of pharmacometrics is the only feasible manner to get optimal drug use for pediatric patients. This is accomplished primarily through the estimation of differing levels of covariate influence on pharmacokinetics and pharmacodynamics, and thus dosing strategies. The impact of covariates on pharmacokinetics and pharmacodynamics is more important in children than in adults because of the large range of influential covariates such as weight and age in children (4, 5). In pediatrics, dosing is most often based on patient size such as body weight, body surface area (BSA), weight bands, or age. The goal of these dosing strategies is to generate near identical exposures (peak concentrations, area under the concentration–time curve, time above a threshold, or trough concentrations) across age–weight groups

adjusted for known PD differences. This goal is complicated by limitations in formulations, where dose may be dictated by solid oral dosage form strengths. There is a need to balance appropriate exposure for the population against overly complex pediatric regimens that are likely to result in abandonment or dosing errors.

39.2   REGULATORY CLIMATE

In 1962, the Kefauver–Harris Drug Amendment was passed to ensure drug efficacy and greater drug safety. For the first time, drug manufacturers were required to prove to the FDA the effectiveness of their products in the treated population before marketing them. In addition, the FDA was given closer control over investigational drug studies, FDA inspectors were granted access to additional company records, and manufacturers had to demonstrate the efficacy of already approved products. There was a conservative climate following the Kefauver–Harris Drug Amendment that led to the avoidance of the study of drugs in pediatric populations. This avoidance led to a lack of pediatric labeling for greater than 90% of drugs in certain pediatric populations.

   The lack of labeling studies for pediatric patients was seen as posing significant health risks to children. Initially, the FDA implemented largely voluntary measures through the Pediatric Rule to encourage the study of drugs in children and to enhance pediatric labeling in the early 1990s. These measures failed. Therefore, in 1997 the Food and Drug Administration Modernization Act (FDAMA) included incentives to study drugs in children through FDAMA and the subsequent Best Pharmaceuticals for Children's Act (BPCA) by adding six months of attached exclusivity to any existing exclusivity. This legislation had a potent effect on increasing drug studies done in children and in a report to Congress it was stated that "the pediatric exclusivity provision has done more to generate clinical studies and useful prescribing information for the pediatric population than any other regulatory or legislative provision to date."

   In 2003 Congress enacted the Pediatric Research Equity Act (PREA) to further promote drug study in the pediatric population. Here the FDA was given the authority to require studies for the registration of a new drug when deemed necessary. Therefore, now the FDA has both a "carrot" (BPCA) and a "stick" (PREA) to encourage the study of drugs in pediatrics.

39.3   OBSTACLES OF PM RESEARCH IN PEDIATRICS

39.3.1   Gaining Permission for Participation

The first hurdle in executing a pediatric pharmacometric (PM) study is ethically obtaining informed consent from the patient. This requires informed consent from at least one and sometimes both parents or a legal guardian. For older pediatric patients, assents are usually required from the study participants themselves. In addition to the complicated logistical issues of getting this consent, if the legitimate goals of the research are not presented well to the parents, it may result in concern that their child or infant is being used as a guinea pig. This compounds the overall

reluctance to allow participation of their child in a study, even if the parents would likely participate in similar studies themselves. Many PM studies require close collaboration between the study site and pediatric subspecialists to develop the trust needed to gain subject enrollment. Even when the parents are agreeable, the potential pediatric participant may decline to participate. In dealing with a vulnerable population, pediatric PM studies must justify design based on some degree of direct benefit. So, as with adult cancer chemotherapy PM studies, where healthy volunteers are not utilized, almost all pediatric PM studies are performed in children with disease or who are at risk for the disease that the drug is being used to treat. Furthermore, the child or parent cannot be coerced or bribed with reimbursement to participate in studies and some pediatric centers' investigational review boards (IRBs) do not allow any monetary or other inducements at all. In situations where direct reimbursement is a barrier to participation, other financial barriers such as parking fees, taxi rides, delivery of supplies, or overnight accommodations can and should be removed.

39.3.2 Study Design Issues

Some study designs are very difficult to utilize or implement in pediatric patients. The need for direct clinical benefit precludes use of "healthy" pediatric volunteers. Crossover designed studies, which are ideal for assessing drug interactions, are extremely difficult to perform and are seldom executed in pediatric PM studies. Even single-dose PM studies, for drugs with very long half-lives, are challenging because is it difficult to get children to participate for a duration long enough to complete the sampling. It is very difficult and disruptive for parents to coordinate their family's schedules to accommodate bringing infants and children back to a PM study center on successive days. Thus, there is a large drop-off in participation when the sampling duration increases from 8 to 12 or 24 hours and beyond. Population approaches to estimation and development of PM models are advantageous in this setting because several samples may be obtained during the first dosing interval and several more may be obtained at steady state. Here the patient can participate for the first several hours after the first dose and return at a convenient time within a flexible time window after steady state is achieved for additional sampling, thus increasing patient retention. It should be further noted that the population approach can often be applied to evaluate drug–drug interactions, thus obviating the need for crossover studies, although randomization is needed to establish causality.

39.3.3 Pediatric PM Studies Are Time Consuming

Almost every aspect of pediatric PM study execution and analysis requires greater resources to perform and the overall study duration is longer than a similar study in adults. Tasks that require greater time include locating study centers, identifying suitable subjects for participation, obtaining informed consent, and coordinating data quality assurance schedules. Rarely can sites coordinate multiple study participants to synchronize their visits together into a weekend or two. Pediatric PM studies collect more variable data than studies in adults; thus, the data clean-up can be time consuming as well. With more than 200-fold range in possible weight (0.5 kg to >100 kg), standard data filters can become problematic. Transcription and

recording errors of weight in pounds rather than kilograms may not be grossly apparent. Rarely is the same single milligram dose used across all subjects. The multiple doses used provide the opportunity for errors in recording and transcribing dose information. In addition, the normal range for many laboratory values is age dependent, so any filter for toxicity grades must account for different values at different ages.

39.3.4   Collinearity

Collinearity refers to a situation where, in the same data set, some of the covariates are highly correlated with others. A high degree of collinearity between covariates important for PM studies has been demonstrated for at least weight, age, body surface area, height, and creatinine clearance (6). This is demonstrated in Figure 39.1. Many standard laboratory test values are age dependent, so that wide covariate PM univariate screens that include serum creatinine, uric acid, alkaline phosphatase, lactate dehydrogenase, bilirubin, or albumin may identify associations that reflect age or size misspecification. There is also collinearity for many PD biomarkers and surrogate endpoints with age; CD4$^+$ cells, blood pressure, and absolute neutrophil count are just a few examples. Formulation differences may also confound age effects. Conditions where disease progression is a prominent feature, especially if it affects drug elimination, will complicate assessment of age and size effects with other covariates. The implications of collinearity in covariates on the model-building process for nonlinear mixed effects modeling have been investigated by

FIGURE 39.1   Collinearity of various demographic variables in pediatric patients. Of special importance are the collinearities of body surface area (BSA), weight (WTKG), height (HTCM), and age.

Bonate (7). He states that covariates showing a high degree of correlation ($r > 0.50$) when included in a model at the same time may indicate that one or both do not improve the model, even when in fact both should be included. Thus, collinearity is a potential source for spurious covariate exclusion in a pediatric PM model. It has to be pointed out that although the Bonate term $r = 0.5$ represents high correlation, $r \geq 0.75$ is generally regarded as high correlation and an r value between 0.5 and less than 0.75 as moderate correlation.

39.3.5 Cost of PM Studies in Pediatrics

In contrast to adult studies, where the cost–benefit ratio is seldom an issue that impacts study design, it can play a significant role in pediatric study design. Single-dose intensive PM studies may cost less than multiple-dose studies and have the additional benefit of rapidly generating data. For drugs where additional exclusivity is being sought from performing a pediatric PM study per the BPCA, there is a race to complete the study and analysis prior to patent expiration. Thus, in some settings the pediatric PM study completion time may be the most important design consideration. However, single-dose studies are not as powerful for determining covariate effects in PM models because of the increased subject homogeneity and smaller number of subjects. Thus, traditional single-dose studies are more likely to oversimplify dosing across age groups. There is the further issue of whether a single-dose study can predict steady-state pharmacometrics, particularly for pharmacokinetics. For drugs where single-dose pharmacokinetics can predict steady-state pharmacokinetics and can be characterized within an 8–12 hour interval, the cost of collecting and analyzing the additional samples needed for standard noncompartmental analysis (NCA) may be less than enrolling additional subjects for a population analysis. However, the knowledge generated is uniformly less with a NCA. Most BPCA-inspired pediatric drug programs also require a larger safety and drug effect study. This may provide the opportunity to rapidly collect data for pediatric population PK/PD evaluations.

39.3.6 Sampling How Much and What

The quantity of sample for adults is rarely an issue, in contrast to pediatric patients where it is almost always an issue. This is especially true when it comes to obtaining blood, plasma, or serum because of the limited quantity that most pediatric patients possess and can safely be collected for purposes of the study. Even in PM studies of larger children, where blood volume is not a safety issue, it is important to minimize unnecessary blood collection volume for both ethical and recruitment considerations. Not only does the total blood volume need to be minimized but also the number of venipunctures. There is much greater subject and parent acceptance of six PK samples drawn over an 8 hour dose interval through placement of an indwelling catheter than three PK samples, each collected by direct venipuncture, drawn at PK optimized collection times over 24 hours. Liquid chromatography tandem mass spectometry (LC MS-MS) is a good method because one can use a small sample size due to the sensitivity and specificity of the assay method. A specific assay may minimize the need for baseline samples, further reducing the quantity of samples needed.

It is a good idea to use an analytical laboratory that specializes in or has extensive experience with pediatric samples. Many adult contract laboratories have default sample volume requirements for standard hematology and chemistry evaluations that are greater than may be needed for the PK sampling of the study. Also, since many clinical chemistry values are age dependent, using a pediatrics-specific laboratory will prevent artificial labeling of many values as outside the normal range, when they are simply outside the normal range for adults but normal for infants or children.

A further recommendation in pediatric PM studies is the use of a local anesthetic, such as Emla cream, whenever possible. Not only will this reduce patient discomfort but it will also ameliorate anxiety from both potential subjects and parents, thus reducing reluctance to study participation. Use of local anesthetics requires planning as application is require well in advance of the venipuncture.

Pediatric PM studies should be done in a pediatrics friendly environment with age-appropriate supplies (toys, games, etc.) and age-specific activities available. The phlebotomy team should be experienced and dedicated to pediatric venipuncture so that prior to obtaining a specimen the patient is properly assessed. One common practice that is helpful in pediatric sample collection is the use of two individuals, one to obtain the sample and one to distract, during the phlebotomy procedure.

A population analysis approach allows for the use of opportunistic blood sampling for pediatric PM studies. Opportunistic samples can be obtained when there is blood, plasma, or serum "left over" from other procedures or processes. This requires use of a very sensitive assay and can be helpful in hospitalized pediatric patients. This is particularly pertinent for infants where sampling is especially limited and for drugs with long half-lives where potentially reduced accuracy in collection times will not adversely affect PM model development.

Another source for PK information to consider is urine, which may be useful for renally eliminated drugs. In newborns, use of pulp derived diapers can provide reasonably accurate collection of urine for determination of renal clearance. Urine collection can provide valuable mass balance information, which may provide additional stability to a PM model. However, use of adhesive bags for urine collection can lead to maceration of the skin and therefore should be employed with caution. Finally, a sampling and study strategy that recognizes that some collection will be incomplete or sample will be lost is best. The quantification of metabolites can also be useful in pediatric PM studies. The metabolite models can actually provide improved information concerning the parent compound as well as providing mechanistic information concerning age effects on drug disposition. When metabolites are active, then improved PD and outcomes models can be developed, increasing model applicability for proposed dosing strategies.

For population pediatric PK studies, Jones et al. (8) have addressed many of the sampling issues by executing Monte Carlo simulations. In these simulation analyses, the authors assumed that a single sample was obtained from each subject in a cross-sectional design. The ability to estimate both a one- and two-compartment model was investigated to evaluate timing and number of samples needed for accurate and precise estimation of parameters including random effects parameters. The timing of these samples was determined using the informative (profile) block randomized design (see Chapter 12). Specifically, the informative times obtained with ADAPT II and the concentration–time profile were divided into three sampling blocks and subjects were randomly sampled within each block. Several sampling schemes were

investigated for both the one- and two-compartment models. Both single-dose and multiple-dose study impacts were also evaluated. For the one-compartment model, the sampling window blocks were 0.06–0.10, 0.1–2.9, and 2.9–5.4 hours. For instance, with sampling scheme A, samples were collected in a proportion of 3:4:3 for the three blocks, while sampling scheme B maintained 20 samples for the simulations with 50 or 100 subjects and 10 samples for the simulations with 20 and 30 subjects at the middle region sampling interval (be reminded that only one sample per subject was observed in the simulation). The detail of the sampling scheme is presented in Tables 39.1 and 39.2. The true parameters were those used in the simulation and therefore the accuracy and precision of the estimates of the parameters could be quantified. Both the degree of bias and precision of estimates relative to "true" values were of interest and were computed. One hundred replicate simulations were done for each scenario and the percent prediction error (%PE) and the percent root mean squared precision error (%RMSE) were estimated for the parameters for bias and precision, respectively. This investigation showed that the sampling strategy impacted the ability to accurately estimate PK parameters. For a one-compartment model, an N of 50 estimated the typical values of CL, V, and between-subject random effect for CL (ω^{CL}) with little bias and good precision.

TABLE 39.1   Sampling Scheme A*

Number of Experimental Units	Number of Samples in Sampling Block 1	Number of Samples in Sampling Block 2	Number of Samples in Sampling Block 3
20	6	8	6
30	9	12	9
50	15	20	15
60	18	24	18
70	21	28	21
80	24	32	24
90	27	36	27
100	30	40	30

*Source: From Jones et al. (8), used with permission from Taylor and Francis Group LLC, www.taylorandfracis.com.

TABLE 39.2   Sampling Scheme B*

Number of Experimental Units	Number of Samples in Sampling Block 1	Number of Samples in Sampling Block 2	Number of Samples in Sampling Block 3
20	5	10	5
30	10	10	10
50	15	20	15
60	20	20	20
70	25	20	25
80	30	20	30
90	35	20	35
100	40	20	40

*Source: From Jones et al. (8), used with permission from Taylor and Francis Group LLC., www.taylorandfracis.com.

The estimate of the between-subject random effect for V (ω^V) lacked precision. Neither sampling scheme (scheme A versus scheme B) appeared to perform better than the other. When the residual variability was increased, bias in the estimation of V increased and the precision of ω^{CL} and ω^V worsened. For the two-compartment model, an N of 80 resulted in accurate estimates of CL, V1, V2, and ω^{CL}, and precise estimates of CL, V2, and ω^{CL}. All estimates of intercompartmental clearance (Q) were inaccurate and imprecise regardless of sample size and the precision of V1 and ω^V were poor regardless of sample size.

This study demonstrated that for typical drugs, when an informative profile block randomized design is used, single sample cross-sectional designs can be constructed to adequately estimate CL and V and their respective variability for populations of 50 subjects for one-compartment drugs and 80 subjects for two-compartment drugs. For drugs with high residual variability, there was worsened accuracy and precision of parameter estimates. For each drug with its own set of PK parameters, the exact informative sample times will be different. If within-subject longitudinal samples are added to a study, the accuracy and precision of all parameters would be expected to improve over the purely cross-sectional studies. Given the ability of these cross-sectional studies to estimate pivotal parameters (CL and V) with accuracy, dosing strategies can be proposed based on their PK parameter estimates. This type of cross-sectional study could be very useful when there are severe restrictions in obtaining samples as is often the case in pediatric studies.

39.4 DIFFERENCES BETWEEN ADULT AND PEDIATRIC PATIENTS

39.4.1 Differences in Pharmacokinetics

39.4.1.1 Absorption
Absorption patterns in pediatric patients are very different from adults and one must be aware of these differences when conducting PM research. For example, absorption can be age dependent because infants have less gastric acid production, which influences agents where acid is needed for absorption such as azole antifungals and atazanvir. Infants also produce less lipase than adults and older children; this enzyme may be necessary for the absorption of some drugs. More recently, it has been noted that active transporters in the gut limit absorption of some compounds. The degree of transporter expression may exhibit age dependency. Although overall GI transit time may be shortened, gastric emptying is delayed, and one may see slow absorption in very young infants (9). Absorption may also be altered due to age-specific formulations with liquid formulations often administered to younger children and infants. For drugs with absorption sensitive to food intake, the ability to fast prior to dosing or take with prescribed food is more difficult to accomplish in infants and younger children.

39.4.1.2 Distribution
Changes in body composition seen throughout infancy and childhood have an impact on the apparent volume of distribution (V). Newborn infants have higher total body water on a L/kg basis than older populations; therefore, drugs distributed to water have higher V values in newborns when compared to older children

or adults. For example, aminoglycoside antibiotics are highly polar and thus are distributed primarily to extracellular fluid and have double the V value in newborns when compared to adults. Body fat content may also be altered especially in preterm infants who have very little fat, reducing V for lipophilic agents.

In neonates, albumin and alpha-1-acid glycoprotein concentrations are low and albumin has a specific "fetal" structure with different binding characteristics compared to "adult" albumin. Infants also have higher concentrations of endogenous compounds that compete for drug binding sites on protein and in tissue. Overall in neonates, highly bound drugs will have a higher free-fraction of drug in plasma, leading to proportionally greater distribution into tissues. Thus, the V will on average be greater. These differences in binding are important to consider in interpreting pediatric PM results. Doses that achieve similar total concentration–time area under the curve (AUC) profiles for adults and infants may be associated with higher free-drug AUCs in infants, possibly resulting in increased toxicity.

39.4.1.3 Metabolism

Metabolism in pediatric patients can be quite different from adults. In the very young infant, drug uptake by the liver is decreased due to reduced transport proteins. The biliary excretion of antibiotics with dual routes of elimination suggests that hepatic transport maturation is even slower than glomerular filtration or renal transport maturation. Overall, mixed function oxidases are present at 30–50% of adult activity, while individual enzymes may be less than 5% of adult activity. In particular, isoenzymes of CYP 2C9 and 1A2 have greatly reduced activity in neonates; however, there is a rapid increase in 2C9 activity in the first weeks of life. After birth, Phase I and II enzymes have a programmed order of expression, which is different for each isoenzyme. Some isoenzymes increase in days, others over weeks, and still others over months.

Infant hepatic mass is two to three times adult mass on a weight basis. Thus, overall enzyme capacity is higher on a weight basis. A study from St. Jude's Children's Hospital correlated clearance with liver size by scan (10). Both hepatic enzyme activity and hepatic blood flow impact liver metabolic clearance. The influence of each component depends on the substrate in question. While liver size may be a surrogate for total liver enzymatic activity, much less is known about development changes in hepatic blood flow. However, given changes in developmental cardiac output, it is likely that hepatic blood flow and clearance of highly extracted drugs correlate more closely with body surface area than body weight.

39.4.1.4 Renal Excretion

At birth, renal blood flow is 12 mL/min and the kidneys receive only 5–6% of cardiac output compared to 15–25% in adults (normalized to body surface area). Glomerular filtration rate (GFR) is directly proportional to gestational age beyond 34 weeks. GFR increases rapidly over the first few weeks of life, with smaller increases throughout the first year of life. Preterm infants have reduced GFR but exhibit more rapid development in the postnatal period than would have occurred in utero if delivered at term. The peripartum use of steroids to promote lung development during preterm labor may expedite maturation of renal function as well.

Tubular secretion is less mature when compared to GFR and increases twofold over the first week of life and tenfold over the first year of life. Renal function

parallels increases in body weight and this collinearity can be problematic when one is attempting to include both covariates in population PK models concurrently. Clinical serum creatinine measurements, which are used to evaluate renal function in adults, have limited precision in estimating renal function in infants and young children due to low granularity in reported results, typically given only with one-tenth of mg/dL precision. Normal serum creatinine values are 0.2 mg/dL, and because of the reporting units this encompasses true concentrations from 0.1501 to 0.2499 mg/dL or a 66% range difference. Additional analysis difficulties can be encountered with renal function development in longitudinal PM studies. While serial assessments at various stages of development can provide information-rich data, the continued growth and maturation result in large within-subject variability in "base" models. This must be recognized and can cause computational difficulties using standard model-building approaches. The first few weeks of life are also characterized by large inter-subject variability resulting from maturation differences among infants.

39.4.2 Differences in Pharmacodynamics

Ongoing growth and development affects PD analyses. When attempting to employ a biomarker or surrogate endpoint, one must be aware that tests or procedures that are easily applied to adults often cannot be used in infants or children. For example, only limited pulmonary function evaluations can be performed in infants because they cannot execute the standard test. Even though newborns feel pain, it is difficult to get assessments of pain in children and pain scales are difficult to compare across age categories. As a substitute for pain scales, physiologic changes such as blood pressure, catecholamine release, and heart rate variability can be employed, but even these are age dependent and may be affected by concomitant therapies used in infants and young children.

Although $CD4^+$ cell count is accepted as a valid surrogate endpoint for HIV disease, its appropriate use as a PD marker in pediatric studies in HIV-infected children is not clear. The $CD4^+$ count is higher in infants compared to older children and does not stabilize until around 5 years of age. Therefore, while "successful" HIV therapy in adults will typically be associated with increases in $CD4^+$ cells, immunologic "success" in infants may be manifested by a smaller drop in $CD4^+$ cells or change in the $CD4^+/CD8^+$ cell ratio.

39.5 COVARIATE IMPACT IN PEDIATRIC PHARMACOMETRICS

39.5.1 Size as a Covariate

Size is a critical element for understanding, analyzing, and applying PM principles to pediatrics. Weight can range more than 200-fold between premature infants and adolescents, and it correlates with age and other factors that may impact drug disposition. However, many physiologic covariates that affect drug clearance such as renal function do not scale directly to weight. Body surface area (BSA) has been found empirically to correlate more closely with the clearance of many drugs rather than weight. This most likely occurs because physiologic processes are slower

in larger individuals than in smaller ones. BSA can be calculated on the basis of several equations (11–13). One problem with BSA is that it is must be estimated from height and weight and is prone to calculation errors.

Allometric scaling of PM parameters is the preferred approach during childhood. A major advantage of allometric size adjustment is that it is a mechanistic approach that is based on dispersion theory (14, 15). This theory suggests that clearance parameters should be scaled by $WT^{0.75}$ and volume parameters $WT^{1.0}$. Use of standardized allometric exponents, besides being mechanistically appropriate, facilitates comparison among models. This approach does not take maturation or age effects into consideration, which must be evaluated separately from the size effects. It also structurally suggests half-life will increase with size and age, which is commonly seen after maturation of CL processes during infancy. Recently, it has been suggested that, based on an evaluation of a large number of xenobiotics across species, the allometric exponent value differs based on route of elimination: 0.67 for clearance is for drugs eliminated mainly by biotransformation and 0.75 is more appropriate for drugs eliminated by the kidney (16). However, the added complexity of a different exponent remains to be justified for assessing size across the age continuum within a species.

It is important to recognize that dosing strategies based on allometry can match AUCs across age groups; however, the peak–trough differences would still exist (with identical dose intervals) and would be greater in smaller (younger) patient populations. This may have importance for drugs whose PD response is linked through a threshold or a peak concentration effect. It is best to include variables of size into PM covariate models prior to incorporating other covariates for useful models.

39.5.2 Age as a Covariate

Age comes in multiple structures for pediatric populations. Infant age may be defined in terms of postnatal and gestational ages, while for adolescents biological age (Tanner scores) may be more appropriate than chronological age. When attempting to evaluate age as an influential covariate, for infants, if data is dense, one should evaluate the impact of postnatal age and gestational age at birth as two separate covariates, thus recognizing that maturation occurs at different rates in in utero versus postnatal environments. Metabolic and excretory functions increase more rapidly following birth than in utero and thus composite measures of infant age such as postconception age are less accurate in this setting. The impact of age can have a nonlinear relationship to PM model parameters and therefore graphics, particularly nonparametric smooths, and generalized additive models should be used to determine the relationship between age and the PM parameter of interest. During an analysis, it must be remembered that age is a dynamic covariate and subjects can have large increases in drug clearance in sequential PK evaluations. While much of the power from population PM analyses in infants can be derived from this mixture of longitudinal with cross-sectional patient evaluations, models without interoccasion variability components will output average parameters within a subject in the posthoc estimates from the base model. Graphics from these outputs will have the true maturation processes blunted and this must be taken into account during the modeling process.

39.5.3 Creatinine Clearance as a Covariate

The impact of creatinine clearance (CLCr) as a predictor variable for pediatric PM parameters is often low even for renally eliminated drugs due to the inclusion of size (height) in the calculation of renal function and its normalization to adult size. Although the original pediatric renal studies by Schwartz (6) show excellent correlation between measured and estimated CLCr from serum creatinine, others have noted these equations as not predictive in some pediatric subpopulations. The poor precision for the clinical assay of serum creatinine partially accounts for this lack of predictability. However, many pediatric PM analyses have demonstrated the reciprocal of serum creatinine is a powerful covariate for predicting the clearance of renally eliminated drugs, even in infants.

39.5.4 Drug Interactions

Drug interaction screens do not determine cause; therefore, caution is needed in interpreting the meaning of a drug interaction covariate effect on pediatric PM parameters. For example, inotropes have been included in models as significant predictors of clearance in infants; however, they likely identify a characteristic of the severity of underlying illness of the subpopulation rather than a true drug interaction. While pediatric patients may receive a different scope of potentially interacting drugs, the expected qualitative effects are typically similar to adults. Enzyme inducers and inhibitors have been identified as covariates for predicting clearance in pediatrics. Some drug interactions only suspected in adults have been documented in pediatric populations such as dapsone and rifabutin. Development may impact the capacity of an enzyme to be induced so there may be important quantitative differences between adults and children. It is also common for standard pediatric dosing to result in a different exposure to the interacting drug than is seen in adults. If the drug interaction is concentration dependent, the altered drug exposure of the inducer or inhibitor may affect the magnitude of the drug interaction in pediatric patients.

39.5.5 Other Covariates

Other covariates that have been identified as important in pediatric PM studies include the level of metabolism in the gastrointestinal tract, ECMO that may be a marker of hypoperfusion, nutrition, and genetics–genomics. Ethnic differences may also exist in pediatric populations while known gender differences in adults are likely absent or greatly reduced.

39.6 POPULATION MODELING IN PEDIATRICS

Population modeling has great utility in pediatric patient populations because less intense and opportunistic sampling can be executed to estimate population parameters such as typical values for clearance. The population approach in pharmacometrics allows for variable dosing regimens, variable sample collections, the use of unbalanced data, the study of a broad spectrum of patients, and a screen for drug–drug interactions, provides estimates of individual drug exposure for explor-

atory analysis of efficacy and toxicity, and allows pooling of data across studies. It is especially useful for incorporating significant covariates into PM models, assessing complex PM models, modeling at steady state, and comparing formulations, and pharmacokinetics–pharmacodynamics–outcomes links models can be generated. The stochastic elements of population models are more accurately estimated than for traditional standard two-stage models (17).

39.6.1 Sampling Strategies in Population Modeling

Of particular interest is the utility of population modeling when sampling is limited, as is often the case in pediatric studies. For example, etoposide toxicity and efficacy have been related to exposure. It was not reasonable to execute an intense sampling PK study in the pediatric population, therefore a limited sampling strategy was proposed and done. In each subject only two samples were collected—one at about 3 hours postdose and another 5.5 hours postdose. The approach was shown to be able to estimate PK parameters that had little bias (18).

Optimal sampling strategies for pediatric population studies can be determined by using Monte Carlo simulation. Here one constructs plausible data sets that would be obtained under several competing study structures by varying study characteristics such as number of subjects, number of samples per subject, missing data, mistimed sampling, and mislabeled sampling. From the basic model, several hundred plausible data sets are generated by the simulation software and then models are estimated for each data set. These competing study structures are then compared for power, efficiency, robustness, and informativeness.

Once the PM models are estimated, then dosing strategies can be proposed based on important covariates. These dosing strategies can be assessed by Monte Carlo strategies prior to a future pediatric patient ever receiving a dose. A drawback to population modeling is that it can be logistically difficult to perform, often a great deal of education is needed at the sites where the data are collected, and quality assurance of the data can be difficult.

39.6.2 When to Incorporate Size in a Population PM Model

A frequently addressed question is whether size should be incorporated into the population model first before other covariates are tested for inclusion, because it is well known that size impacts pharmacokinetics especially in pediatric patients. A further and related question is: Should one fix the allometric exponent or should one estimate the exponent? The advantage of estimating the allometric exponent is that a statistically superior model may be found. However, this is at the expense of two degrees of freedom from studies where the total information may be limited. Fixing the allometric exponents requires making additional assumptions about the relationship of size to the parameter but it increases the utility of the resulting model when comparing to other models. In addition, situations where the allometric exponent deviates significantly from the expected are often a reflection of developmental differences among the study population, with changing exponent values depending on the age group included in the analysis.

One problem with allometric scaling is that it is difficult for those not familiar with the concept to grasp the meaning of clearance as $L/h \cdot kg^{0.75}$ and how it may

affect dosing. Some advocate presenting the parameters at the median or mean weight of the study population. While this is preferable to the raw exponents, the concept of normalizing clearance to size is so ingrained that many pediatric clinicians will then divide the value by the typical weight clearance in L/h·kg. To prevent this mistake, one should also present the model predicted clearance for the smallest and largest subjects (this can be done graphically) to provide a clearer representation of clearance across the population.

39.7   CLINICAL TRIAL SIMULATION

One of the most potent applications of pharmacometrics is the informative construction of clinical trials by using clinical trial simulation (CTS). Population PM models are of great value when used in CTS because estimates of typical parameters along with parameter variability can be incorporated. There are three basic types of models needed to execute a CTS: an input–output model, a covariate model, and an execution model. These are described in detail in Chapter 34 of this book. Clinical trial simulation can improve pediatric study structure by examining the impact of many important factors such as dropouts, choosing varying endpoints, and deviations from protocol. Pediatric PM models find great utility when applied to CTS.

39.8   AN INFORMATIVE EXAMPLE

We present a pediatric population PK (PPK) model development example to illustrate the impact that the model development approach to scaling parameters by size can have on pediatric PPK analyses; a typical pediatric study is included. It is intuitive that patient size will affect PK parameters such as clearance, apparent volume, and intercompartmental clearance; and that the range of patient size in most pediatric PPK data sets is large. Thus, it is expected that in most pediatric PPK studies subject size will affect multiple PK parameters. However, because there are complex interactions between covariates and parameters in pediatric populations, there are also intrinsic pitfalls of stepwise forward covariate inclusion. Selection of significant covariates via backward elimination has appeal in nonlinear model building; however, it requires knowledge of the relationship between the covariate and model parameters (linear vs. nonlinear impact) and can encounter numerical difficulties with complex models and limited volume of data often available from pediatric studies. Thus, there is a need for PK analysis of pediatric data to treat size as a "special" covariate. Specifically, it is important to incorporate it into the model, in a mechanistically appropriate manner, prior to evaluations of other covariates.

The current example is drawn from results of a PK study that was designed to evaluate the pharmacokinetics of cyclosporine in stable pediatric transplant patients receiving chronic oral dosing. Since many of the subjects had evaluations from two separate formulations, a secondary objective of the study was to evaluate the relative absorption characteristics of the two formulations. The study included 32 children and adolescents, a typical size for a pediatric Phase 1–2 study. This modest number of subjects in a pediatric PK study is common but reduces the power to

include extensive covariates in a final PK model. Therefore, it is essential to limit the scope of covariates that will be evaluated to those related to study objectives and those that are expected to impact dosing.

In this study, there was relatively intensive cyclosporine PK samplings: eight samples per subject were collected over a dose interval of the primary formulation. Many subjects had an additional, limited, three-sample PK evaluation performed a few months after the primary PK evaluation. In addition, a subset of subjects had previously participated in another separate cyclosporine PK study utilizing a different formulation. Their PK data from this prior study were included in the analysis. The average subject age in this study was 13.7 years with a range of 3–21 years. Like most pediatric trials, there was a large range in overall subject size that reflected true differences in absolute liver and kidney organ mass and function. The weight range encompassed nearly a tenfold range (from 12.2 to 121 kg) and BSA averaged 1.32 m^2 (range 0.53–2.39 m^2). This study did not represent the entire pediatric continuum, as no infants below 3 years of age were included. The PK analysis issues related to size would have been magnified further if infants had been included, because not only the range of size would have been expanded but also the diversity of hepatic and renal elimination maturation issues would have been encountered. For the development of this model, the a priori level of decrease in the minimized objective function (MOF) for retaining covariates in the model was set at 6.6 ($p < 0.01$) for a single degree of freedom in hierarchical models. A two-compartment model greatly improved the fit of the data as indicated by a greater than 100 point decrease in the MOF. The first-order conditional estimates algorithm was utilized after a log transformation of concentration data.

The most commonly used approach to population PK model building, developed in adults, starts with a base model that contains no covariates for PK parameters and adds covariates one at a time, assessing the impact on the model by changes in the MOF. Weight or other metrics of subject size are added with size covariates to each PK parameter separately. This approach is not optimal for pediatric PK modeling, and utilizing it in this cyclosporine example resulted in no inclusion of size covariates for any PK parameter in the final model. To illustrate this point, models were developed starting with a base model devoid of size covariates and adding individual covariates of BSA or weight to CL, V2, or V3 one at a time. From this approach, no covariates resulted in a reduction in the MOF by 6.6 and only one decreased the MOF by more than 4, weight added to V3 (decreased the MOF by 4.5). These results occurred despite the clear post hoc graphical suggestion of size effects on CL (see Figure 39.2). However, when size covariates were applied to CL, V2, V3, and intercompartmental clearance (Q) simultaneously, the objective function was significantly reduced by over 25 ($p < 0.001$ on 4 df). The univariate screening approach would have become even more confusing if many covariates were evaluated in this step due to multiple collinearities between size, age, and serum creatinine. Age as a covariate had a similar impact to WT and BSA on PK parameters and in some instances was the most powerful covariate.

Although simultaneous inclusion of size on all parameters improved the model, some of the size covariates did not remain statistically significant for all parameters if evaluated by stepwise backward elimination. While mechanistically this is not a plausible reflection of the true nature of cyclosporine's pediatric PK disposition, it is not totally unexpected that all of these covariates do not all reach "statistical

FIGURE 39.2 Relationship between BSA and total clearance demonstrated by graphical presentation from post hoc estimates of a base model without size. Note that BSA was not a significant covariate in the univariate screen when included in isolation size effects on parameters despite the obvious relationship between BSA and clearance.

significance" based on the modest study size and the limited information on some PK parameters from the sampling design. Removing individual size effects would result in a model with peculiar characteristics. It makes mechanistic sense that all of the PK parameters of CL, V2, V3, and Q would be expected to increase with increasing size; that is, the smallest patient in this study (12.2 kg) would not be expected to have the same value for any of these four parameters as the largest patient (121.0 kg). While a model developed through pruning size effects from some PK parameters would be statistically correct, the usefulness would be severely limited. The complexity of a two-compartment model for this data, although clearly justified by the goodness-of-fit plots and MOF (MOF for one-compartment model was 100 points more than for a two-compartment model), may have limited the ability of the univariate screening to detect size covariates. However, the one-by-one addition of size covariates to PK parameters to a one-compartment model also failed to decrease the objective function by 4 in any permutation.

A preferable approach that was used in this analysis was to start with allometrically scaled parameters, clearance and volume terms scaled by $WT^{0.75}$ and $WT^{1.0}$, respectively, before assessment of other covariates. The underlying assumption that these size relationships exist is plausible and results in a base model that has the same degrees of freedom as a base model without the size assumption. In the cyclosporine analysis, allometric scaling accounted for much of the apparent pediatric PK "age" effects. However, even after scaling CL by $WT^{0.75}$ in the final model, children less than 12 years of age were associated with a 26% higher clearance. Formulation had a similar magnitude impact on F (24%). Different formulation F was associated with a drop in the objective function of 12. The final model included the covariance estimation between CL and V2 as well as interoccasion variability on F. The final model is presented in Table 39.3.

TABLE 39.3   Complete Description of the Final Irreducible Model

Parameter	Value[a]	SEE
CL	$1.86 \cdot WT^{0.75} \cdot AGE12$	0.22 (AGE12: 0.09)
V2	$1.58 \cdot WT$	0.35
KA	0.65	0.07
F1	$1 + 0.24 \cdot FORM1$	0.13
Q	$1.48 \cdot WT^{0.75}$	0.21
V3	$7.75 \cdot WT$	2.41
Intersubject Variability		
CL	23%	13%
V2	65%	36%
IOV—F1	31%	18%
Residual Error	33%	14%

[a] Where AGE12 = 1.27 if ≤12 or 1 if >12 and where FORM1 = 1 if formulation 1 and 0 for formulation 2.

The one-by-one addition of all covariates to PK parameters in the model, treating size just like any other covariate, would have resulted in a model that would not make physiologic–pharmacologic sense. It is often difficult to tease out independent effects of the covariates from pediatric data. Incorporation of prior knowledge of the pediatric physiology, development, disease presentation, and disease progression is often needed to provide important guidance in model development. As one would expect, age and size are moderately correlated in this data set, the R^2 was 0.57 between age and weight and 0.68 between age and BSA. If one were also to include time since transplant, serum creatinine, or other age-dependent laboratory measures that vary with size or age, a typical univariate screen could select any of these potential associations over weight, allometric scaling, or BSA. Thus, a model suggesting that clearance and absolute dose (in mg) be solely a function of time since organ received could be the best statistical model yet a harmful model in application outside the study population. It would grossly underdose a 17 year old with new transplant and possibly overdose a very young transplant recipient. In the standard one-by-one covariate addition model-building paradigm, most of these covariates will drop out at the multivariate step, but "cluttering" this stage with many confounded variables can greatly affect the approach taken in testing the covariate in a multivariate step (categorical, linear, or nonlinear) and one may miss the covariate that has a true causal influence. The inclusion of many covariates due to collinearities with size can also change the order of evaluation in the forward selection process or result in computation difficulties due to the large number of factors if one assesses independence using a backward elimination selection approach. Size, weight, or BSA is intuitively related to CL and V and therefore should be included as prior knowledge in all pediatric population PK modeling exercises.

39.9   SUMMARY

Population PM methodologies represent a powerful approach to generating clinical pharmacology data in infants and children. In the last 15 years there has

been a great increase in the use of this method to analyze PK data. Its potential role in pediatric PD model and disease model development remains to be cultivated. In pediatric PM trials, it is important that they are designed to ask the right questions. Pediatric PM data are often limited and assumptions may be required for the analysis. Thus, it is essential that pediatric expertise be sought to assist in study design and analysis. Useful models should be mechanistically relevant and, at a minimum, should account for size and assess developmental changes to be useful. With the increased performance of pediatric population PM studies, opportunities exist to learn more about pharmacologic ontogeny and build better developmental models. These models may ultimately enhance the safety and effective use of drugs in this important population.

REFERENCES

1. J. L. Hiller, G. I. Benda, M. Rahatzad, J. R. Allen, D. H. Culver, C. V. Carlson, and J. W. Reynolds, Benzyl alcohol toxicity: Impact on mortality on intraventrucular hemmorrhage among very low birth weight infants. *Pediatrics* **77**:500–506 (1986).

2. T. M. O'Shea, J. M. Kothadia, K. L. Klinepeter, D. J. Goldstein, B. G. Jackson, R. G. Weaver, and R. G. Dillard, Randomized placebo-controlled trial of a 42 day tapering course of dexamethasone to reduce the duration of ventilator dependency in very low birth weight infants: outcome of study participants at 1-year adjusted age. *Pediatrics* **104**:15–21 (1999).

3. Department of Health and Human Services, *Guidance for Industry: General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products*. U.S. Food and Drug Administration, Rockville, MD, 1998.

4. P. Rajagopalan and M. R. Gastonaguay, Population pharmacokinetics of ciprofloxacin in pediatric patients. *J Clin Pharamacol* **43**:698–710 (2003).

5. E. Chatelut, A. V. Boddy, and B. Peng, Population pharmacokinetics of carboplatin in children. *Clin Pharmacol Ther* **59**:436–443 (1996).

6. G. J. Schwartz, G. B. Haycock, C. M. Edelmann, and A. Spitzer, A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine. *Pediatrics* **58**:259–263 (1976).

7. P. Bonate, The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharm Res* **16**:709–717 (1999).

8. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional population pharmacokinetic studies: Implications for pediatric and animal studies. *Clin Research Reg Affairs* **13**:133–165 (1996).

9. G. Kearns, P. K. Robinson, J. Wilson, D. Wilson-Costello, G. R. Knight, R. M. Ward, and J. van den Anker, Cisapride disposition in neonates and infants: in vivo reflection of cytochrome P450 3A4 ontogeny. *Clin Pharmacol Therap* **74**:312–325 (2003).

10. D. J. Murry, W. R. Crom, W. E. Reddick, R. Bhargava, and W. E. Evans, Liver volume as a determinant of drug clearance in children and adolescents. *Drug Metab Dispos* **23**:1110–1116 (1995).

11. D. Dubois and E. Dubois, A formula to estimate the approximate surface area if height and weight be known. *Arch Int Med* 17863–17871 (1916).

12. E. A. Gehan and S. L. George, Estimation of human body surface area from height and weight. *Cancer* **54**:225–235 (1970).

13. R. D. Mosteller, Simplified calculation of body-surface area. *N Eng J Med* **317**:1098 (1987).

14. G. B. West, J. H. Brown, and B. J. Enquist, The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**:1677–1679 (1999).

15. G. B. West, J. H. Brown, and B. J. Enquist, A general model for the origin of allometric scaling laws in biology. *Science* **276**:122–126 (1997).

16. T. M. Hu and W. L. Hayton, Allometric scaling of xenobiotic clearance: uncertainty versus universality. *AAPS Pharm Sci* **3**:E29 (2001).

17. L. B. Sheiner and S. L. Beal, Pharmacokinetic parameter estimates from several least squares procedures: superiority of extended least squares. *J Pharmacokinet Biopharm* **13**(2):185–201 (1985).

18. J. C. Panetta, M. Wilkinson, C. H. Pui, and M. V. Relling, Limited and optimal sampling strategies for etoposide and etoposide catechol in children with leukemia. *J Pharmacokinet Pharmacodyn* **29**(2):171–188 (2002).

Pharmacometric Methods for Assessing Drug-Induced QT and QTc Prolongations for Non-antiarrhythmic Drugs

HE SUN

40.1 INTRODUCTION

The electrocardiogram has been intensely studied. Figure 40.1 demonstrates the waves of a heartbeat as recorded on the electrocardiogram (ECG) rhythm strip. They are labeled, according to well accepted practice, with the letters P, Q, R, S, and T. This chapter primarily addresses the QT segment, as recorded on the ECG rhythm strip, which includes the time interval (measured in milliseconds, ms) from the beginning of ventricular depolarization, the Q wave, to the end of the T wave, at which point cardiac repolarization is complete. QT prolongation refers to lengthening of a normal QT interval.

While the extent of QT prolongation is acknowledged as an imperfect biomarker for proarrhythmic risk, there is a quantitative relationship between QT prolongation and the risk of torsades de pointes (TdP), especially for drugs that cause substantial prolongation of the QT interval (1). Because of the QT's inverse relationship to heart rate (HR), the measured QT interval is routinely corrected by various formulas that relate the QT to the HR, known as the QTc interval. Although it is not clear whether arrhythmia development is more closely related to an increase in the absolute QT interval or QTc, most drugs that have caused TdP clearly increase both the absolute QT and the QTc. In pharmacometric analysis, QTc is used as the biomarker of choice for drug-induced QT change assessment.

Several new developments on QT/QTc prolongation assessment have occurred recently. The most important includes the publication of the new International Conference of Harmonisation (ICH), Step 4 guidance, issued on May 12, 2005 entitled *The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*. It contains relevant information obtained from the ICH Steering Committee, as well as information discussed at the October 2003 Food and Drug Administration (FDA) and Drug Information Association (DIA)

FIGURE 40.1 ECG rhythm strip. The extension of the QT interval is termed QT prolongation. The ending point of the T wave sometimes is difficult to determine, which contributes to the variability of QT measurements.

Meeting, "ECGs in Clinical Trials: The New Regulatory Realities," and provides recommendations to sponsors concerning the design, conduct, analysis, and interpretation of clinical studies to assess the potential of a drug to delay cardiac repolarization. It includes testing the effects of new agents on the QT/QTc interval and the collection of cardiovascular adverse events (AEs). Readers are encouraged to access this document for detailed information.

One key point in the E14 guidance for a pharmacometrician who is conducting QTc data analysis is that "a negative thorough QT/QTc study is one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval excludes 10ms. This definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QT/QTc interval is not greater than around 5ms. When the largest time-matched difference exceeds the threshold, the study is termed 'positive QT trial.' A positive study influences the evaluations carried out during later stages of drug development, but does not imply that the drug is pro-arrhythmic." Additional statistical guidance can also be obtained from the PhRMA Working Group paper released in 2003 (2).

The FDA will consider substantial QT/QTc interval prolongation, with or without documented arrhythmias, as grounds for nonapproval or discontinuation of clinical development, or require the sponsor to include relevant information in the product label. If it is a feature shared by other drugs of the therapeutic class, it may require a study to compare the extent and incidence of any QT/QTc interval prolongation effects to other drugs in the same class with concurrent positive control groups in the trials (see later discussions on the use of positive control). Documented experiences from the literature where CYP450 metabolism resulted in QTc prolongation and torsades de pointes have prompted regulatory actions such as label warnings or market withdrawal.

ICH, FDA, and industry will benefit by working together in order to clarify outstanding issues on QT/QTc interval prolongation. Additional QT/QTc interval prolongation guidances or methodologies will be issued as ICH, FDA, and industry collaborate on this topic. As the FDA continues to provide more guidance and additional data from thorough QT/QTc clinical studies are submitted to the FDA, additional modifications of new product labels can be expected.

40.2   CORRECTION OF THE QT INTERVAL FOR HEART RATE

Heart rate or RR interval (defined as 60 divided by the heart rate) correction plays a very important role in the analysis of QT/QTc data. Baseline corrections depend heavily on the clinical assumptions that the baseline data represent the subjects' physiological condition in drug-free conditions, and that the QT/HR relationship stays the same before and after drug administration, whether or not the heart rate changes. With this assumption, QT values are corrected for HR in order to sum- marize and compare QTc values across different subjects, trials, or conditions.

The objective of correcting the QT interval for HR or RR is to obtain a corrected QT interval that is statistically independent of the HR or RR interval. Figure 40.2 shows the dependence of QT on HR. In order to eliminate the dependence of QT on heart rate, numerous HR or RR correction formulas have been proposed in the ECG literature, reflecting the variety of statistical models that have been fit to the data. The reader should be aware that there is no best QT interval correction method for heart rate, but there are some practical methods. The most popular corrections are the Bazett (3) and Fridericia (4) formulas. Both are based on the simple power model $QTc = QT/RR^b$; that is, calculation of the QTc is equal to the observed QT in milliseconds divided by the term of a root of the RR interval in milliseconds.

Bazett's method uses $b = 0.5$, and Fridericia's correction uses $b = 0.333$. Both can produce similar or different QTc intervals for the same QT, depending on the HR value. For RR values less than 1 second (i.e., HR greater than 60 bpm), the square root function is smaller than the cube root function and the Bazett-corrected QT interval (QTcB) will be larger than the Fridericia-corrected QT interval (QTcF). Thus, at high heart rates, QTcB is much larger than QTcF and the Bazett formula may "overcorrect" QT interval. Thus, when a drug increases the heart rate substan- tially but does not truly prolong the QT interval, the use of Bazett's formula can inflate the probability of concluding a positive QT/QTc signal when such a signal

FIGURE 40.2   Plot of baseline QT versus heart rate. Two methods of QT interval measure- ment are presented—the manual read and machine read. QT interval is clearly depended on the heart rate. Also, machine read QT interval is longer than manual read QT interval.

does not exist. Similarly, Fridericia's formula is said to "somewhat undercorrect" at low heart rates. In most QTc data submitted in a New Drug Application (NDA), with the same data set, the QTc prolongation assessment endpoint of maximum mean QTc change is always larger for QTcB than for QTcF. This mainly is due to the fact that the mean HR is normally greater than 60 bpm.

In fact, Bazett's and Fridericia's formulas are just specific cases for the data available. In a new study, linear or log-linear models can also be used to derive an empirical population- or subject-specific correction based on the observed QT/RR baseline (predrug) data, including ECGs from all subjects in the study. For example, a pharmacometrician can fit $\log(QT)$ versus $b \log(RR)$ to drug-free data from all patients in all periods to obtain an estimate of b, then calculate QTc by applying $QTc = QT/RR^b$ to all treatment arm/period data. This will generate population-corrected QTc data (QTcP). Similarly, if $\log(QT)$ versus $b \log(RR)$ is fit to baseline data from individual subjects for individual estimates of b, then calculating QTc by applying $QTc = QT/RR^b$ to the individual subject at treatment will generate the subject-specific corrected or individual corrected QTc (QTcI). In most cases, population b values are in the range of 0.22–0.6, and individual b values range from 0.1 to 0.8.

Subject-specific corrections may be preferable to population-based correction formulas in studies with multiple ECG recordings per subject at baseline (5), but their effective use depends on an adequate range of heart rates in the baseline data for each subject. If subjects are in resting condition during the experiment, their heart rates do not usually vary much. Therefore, an individually derived correction based on a narrow range of heart rate pretreatment may not be the most accurate correction available to correct for QT measurements and can lead to false conclusions. Also, considering the large interbeat variability for QT interval (see later for variability), the amount of baseline data needed to generate QTcI is substantially large. These problems clearly have prevented the broad use of the individual correction method and are why QTcI data in NDA submissions were rarely seen.

An important reminder is that the effectiveness of any correction formula for a particular set of QT/RR data from a population or an individual should be examined graphically. If the correction is adequate, QTc will be statistically independent of HR/RR. See the example plots in Figure 40.3 for examining correction effect on QT intervals. In the four panels of plots for different correction methods applied to this data, QTcF provides the best HR correction since the slope of QTcF versus HR is most close to zero (i.e., QTcF is independent of HR). However, a slope of zero in a population may not be enough for an adequate correction.

No single correction formula will work for every data set, and therefore understanding the limitations of each correction is critical (6). For most populations, Fridericia's formula is generally simple to apply and provides an acceptable correction and is the one the FDA normally uses in new drug evaluations. See ICH 14 guidance (7) for further discussion. Although the Fridericia correction is often designated as the primary correction, it is prudent to present the results using several corrections. The FDA encourages sponsors to include population and individual empirical corrections derived from baseline data, in addition to the Fridericia method. When statistical results differ because various corrections were applied to the same data set, which is often seen, discrepancies should be explained and can usually be traced to the range of heart rates prior to and after drug administration, a change

FIGURE 40.3 Heart rate corrected QT versus heart rate. Four methods are applied to the same data set. From the resulting slopes, QTcI (individual correction method) provided the best correction. The Bazett method yields the worst correction.

in heart rate, a change in the QT/RR relationship, or differences between subjects in individual QT/RR relationships. It is recommended that pre- and postdose RR intervals (or heart rates) as well as uncorrected QT intervals also be examined and analyzed in order to understand the relationships between the variables.

40.3 DATA ANALYSIS CONSIDERATIONS IN STUDY DESIGN

40.3.1 Study Type and Number of Subjects

The ICH 14 guidance has clearly indicated that time-matched baseline and placebo correction should be used in QT prolongation assessment. Although the choice of a crossover or parallel group, single or multiple dose, study design is usually determined by the primary objectives of the study, the pharmacokinetic (PK) and pharmacodynamic (PD) properties of the compound (e.g., half-life of the drug, expected margin, delay, or accumulation of QTc prolongation), and the current knowledge of the drug's safety and tolerance profile, a single-dose crossover design is easier for time-matched correction for each individual subject. Also, as normal healthy volunteers are generally to be employed in the design of thorough QTc

trials (TQT), the application of crossover designs is of better utility in improving the accuracy and precision of the relevant prolongation measurements.

Whenever possible and appropriate, the four basic components—the baseline, placebo, positive control, and multiple dose levels of the drug—will assist in evaluating QT/QTc prolongation. Here, baseline data is used for HR correction; placebo control should be employed as QTc is known to show diurnal variation and in order to provide a reference basis for an assessment and potential claim of no-effect; a concurrent positive control group or arm is strongly encouraged and most of time is required; and finally, one or more dose levels that represent the worst scenarios of clinical drug exposure should be included. The most common procedure of a TQT trial has this order: baseline day for all subjects, positive control treatment for all subjects, then crossover design for placebo and treatment drug, or crossover design for placebo, and two or more dose levels of the tested drug. There are other designs that have a separate parallel group for placebo control, or sequential parallel designs, or dose escalation design. Table 40.1 presents a list of various designs seen for recent non-antiarrhythmic drugs.

Due to the large inter- and intrasubject variability in QT/QTc (see later for variability discussion), the sample size for crossover study design would typically be about 60–80 subjects (typically 65 subjects) exposed per treatment. This is to ensure

TABLE 40.1   List of Various Study Designs of Nine Recently Approved Non-antiarrhythmic Drugs

Drug	Doses Used[a]	Trial Design[b]	Number of Subjects	Baseline Collection Period (Data Were Used for Heart Rate Corrections)	Number of Replicate QT Measures Per Time Point
Drug 1	SD	XO	68	On day 1: multiple points until 8h postdosing	1
Drug 2	SD	XO	44	On day 1: only 1h before dosing	3
Drug 3	SD	XO	58	On day 1: at 30, 15, and 0min before dosing	6
Drug 4	MD	PL	40	On day 1: multiple points until 8h postdosing	3
Drug 5	MD	ESC	85	On day 1: multiple points until 24h postdosing	3
Drug 6	MD	ESC	25	On day 1: multiple points until 24h postdosing	3
Drug 7.1	SD	XO	48	On day 1: multiple points until 24h postdosing	1
Drug 7.2	SD	XO	61	Right before dosing	1
Drug 8	SD	XO	90	On day 1: multiple points until 12h postdosing	10
Drug 9	MD	PL	76	On day 1: multiple points until 4h postdosing	3

[a] SD, single dose; MD, multiple dose.
[b] XO, crossover design; ESC, dose escalation design; PL, parallel design.

adequate study power Thus, for parallel study with four treatment arms, 260 subjects would be required, whereas for crossover design, only 65 volunteers (allow for a few dropouts) would need to be exposed to each treatment and washout period. Again, variability is the key determinant of study sample size.

40.3.2 Number and Timing of ECG Recordings

One of the most significant clarifications that evolved from ICH E14 was the move from time-averaged to time-matched analysis (7). A problem with the time-averaged analysis is that the maximum effect of drug on QT interval is diluted, and therefore in the absence of a separate analysis of maximum mean QT change, a false-negative conclusion may be obtained. The advantage of the time-matched analysis is that each time point on treatment (active drug, placebo, or positive control) is compared with the baseline values for the corresponding time point (i.e., to calculate the time-matched ΔQTc). The ΔQTc value at each time point is calculated using the average of the replicated QTc values taken at that time point for each individual. This is the change of QTc from baseline at that time. The parameter of interest is the difference between the ΔQTc on the active treatment (or positive control) and the placebo at the same time (i.e., the ΔΔQTc). When the same baseline data is used, this is equivalent to subtracting the QTc value of placebo from the active treatment or positive control. This approach requires the measurement at baseline and during treatment (placebo, active control) to be the same. The potential problem with this requirement is that if there is a missing value at baseline, or placebo, the data at treatment arm will be useless. No imputation methods to adjust such a missing value have been used or published so far.

The optimal number and timing of ECG recordings in a TQT clinical trial is an area of active research and depends mainly on the endpoint being evaluated, the PK properties of the drug, and the stage of development. The optimal timing to cover a range of concentrations for PK/PD analyses is important. In general, it is recommended to use the time schedule that was normally used for the PK study, or to seek statistical guidance to select the optimal number and timing of ECGs for the objectives of a study, taking into consideration the subject population, endpoint, statistical model, sample size, and cost effectiveness. Normally, 12 time points over the concentration–time profile, in which several points are near the time of maximum drug concentration, are recommended.

Replicated ECGs are needed to avoid the bias in outcome parameters. Malik and Camm (8) recommend that it would be "worthwhile to consider recording 3 to 5 replicate ECGs at each time point within a 2 to 5 minute period." This is because QT interval is not an absolute constant and it is measured with error. The clinical assumption is that even under stable conditions, an individual's true QT/QTc interval can vary largely within a minute. How much of the variability is under stable conditions depends on the natural biological variability and the measurement error. The number of replicates and the number of subjects per trial jointly influence the outcome of a TQT trial. Sun et al. (9) evaluated QT interval variabilities in its chaotic (beat-to-beat), circadian (across a few minutes/hours), and occasion (across days/weeks) domains and provided a reference for clinical trial designs for TQT prolongation assessment. In the study, QTcB, QTcF, and QTcI data from 57 normal young healthy male subjects, with 6 replicated QTs

at each of 3 protocol time points within 10 minute intervals, for 6 cycle days one week apart were used. Standard deviations (SDs) were indices for variability. The impact of inter- and intrasubject variabilities on sample size was estimated statistically and/or via simulations. The overall mean QTcB, QTcF, and QTcI were 391 ± 15.35, 386 ± 14.62, and 381 ± 14.72, respectively, showing that the intersubject variability is independent of correction methods. With QTcF, the average chaotic SD is 14.71 ms (range 9–41). The lowest circadian SD in 20 min, based on the means of 6 replicates per time, was 6.69 ms (range 2.78–11.8). The lowest between-period SD, using the means of all 18 measures for each day, is 7.1 ms (range 1.52–25). These components of variability increased nonlinearly as the number of replicated QTs decreased. Data are presented in Table 40.2. Sun et al. (10) conducted a bootstrap resampling simulation trial. From 80 subjects, 10, 20, 30, 40, or 60 subjects were randomly drawn from drug-free baseline data. Each subject provided two sets of 1, 2, 3, 4, 5, or 6 replicated QT values 10 min apart. The $\Delta\Delta$QTc values were calculated. Since the true $\Delta\Delta$QTc was best believed to be zero, any trials with $\Delta\Delta$QTc greater than zero were considered false-positive results. Figure 40.4 shows the large intersubject variability within a few minutes, and in 30 minute periods. Figure 40.5 shows the distribution of the SD and the high–low limits of 1000 bootstrap resampling analysis results.

When the primary objective of a study is to estimate change in QTc at a specific point in time, say, T_{max}, the use of replicate ECGs can reduce uncertainty. Agin et al. (11) report that for time-matched or within-day changes from baseline, the use of triplicate ECGs instead of single ECGs reduced the within-subject standard deviation as estimated by two different methods: the standard deviation of the observed changes from baseline was 14.7 to 9.2 ms and from 13.5 to 8.1 ms for model-based estimates. Using replicates in this case can substantially decrease the sample size necessary to estimate a response with a desirable precision or test a hypothesis with a prespecified power. With the known large intrasubject chaotic, circadian, and day-to-day variability in QT intervals, Sun et al. (10) independently investigated the frequency of QT replications and the corresponding limit of QT prolongation assessment. Baseline only QT data were obtained from several prospectively designed TQT trials in normal young healthy male or female subjects ($N \geq 40$ in all trials), with up to 6 replicated QT values at multiple time points of a day (intervals ranged from a few minutes to a few hours) for 2–6 treatment periods. One to six jackknife randomly resampled QT values were drawn from each of the

TABLE 40.2   Intersubject Variability as a Function of the Number of Replicate QT Measurements

Number of Replicates	Mean of Within-Subject SDs for All 57 Subjects	SD of the Within-Subject SDs for All 57 Subjects	Ranges of SD for All 57 Subjects
1	6.82	4.31	1.52–12.5
2	5.04	2.96	0.5–11.16
3	4.06	2.3	0.19–10.9
4	3.91	2.1	1.0–12.25
5	3.79	1.98	0.8–9.8
6	3.08	1.85	0.44–8.35

FIGURE 40.4 Example of large intersubject variability in QTcF for an individual. ECGs were recorded six times, 1 minute apart, and repeated three times in 30 min and for six different time periods one week apart.

FIGURE 40.5 Relationship between the number of replicates per time point versus potential experimental bias.

time points and the means, SDs, and changes of means from time to time were determined. The inter- and/or intrasubject QT variabilities and the lowest limit of reliable QT replicates were estimated from 1000 jackknife data sets. Results show that with 1–6 replicates per time point, the intrasubject SD nonlinearly decreased from 16 to 7.82 ms, the time to time QT changes were never zero, and the natural existing QT change ranged from 9.51 to 4.35 ms and maximized from 12 to 49 ms. The intersubject variability is less affected by the frequency of sample replicates. It was discovered that increasing the number of replicates to more than 3 is not needed, as shown in Figure 40.4.

40.3.3 Baseline Days

Since time-matched analysis will be used, the baseline ECGs should be recorded at the same time of day as ECGs collected during active treatment. This will provide insight into diurnal and food effect on the QT/QTc interval. It is important to note that due to the high intrasubject variability, the replicated baseline alone may show false-positive or false-negative QTc changes. Figure 40.6 is an analysis result by Lee et al. (12) with data from the same group of subjects who provided drug–free QT values on consecutive days.

40.3.4 Choice of Endpoint

Many endpoints can be considered for analysis of QT/QTc intervals. The choice should be dictated by the primary objective of the study as well as by the different

FIGURE 40.6 Example to show the diurnal and food effect on the QT/QTc interval over an 18h period.

analyses performed to achieve that objective. It is advisable to mention that other approaches have been used to derive these endpoints. Among the various end-points, the FDA/ICH concept paper gives greater weight to time-matched QT/QTc interval changes and the maximum mean change (and 90% CI) in the QT/QTc interval for the analysis of central tendency.

It is important to point out that the commonly used upper threshold of CI <10 ms as the minimal difference of clinical significance is applicable to the majority of drugs. There are two different summary statistics.

40.3.4.1   Maximum Mean Change
To determine maximum mean change, the means of replicated QTc are first calculated. Then the change from the corresponding baseline is determined (i.e., ΔQTc) for each subject at each matched time point for each treatment. The change from placebo treatment is calculated for each subject at each matched time point (i.e., ΔΔQTc). Finally, the QTc values across all subjects are summarized for means at each time point. The maximum value of the means and the 90% CI for this maximum mean value are determined as the statistic endpoint.

40.3.4.2   Mean Maximum Change
To determine mean maximum change, the means of replicated QTc are first calculated. Then the change from the corresponding baseline is determined (i.e., ΔQTc) for each subject at each matched time point for each treatment. The change from placebo treatment is calculated for each subject at each matched time point (i.e., ΔΔQTc). Then the maximum value of ΔΔQTc for each subject (which can occur at different time points postdose for different subjects) is selected, and finally, the mean and 90% CI of these maximum values for each subject are calculated. The mean maximum change of QT normally lacks statistical property and is greater than the maximum mean change.

40.3.5   Interpretations of Mean Findings with Positive and Placebo Controls

The specific strategy for analysis of QT/QTc data is dependent on the objectives and design of the study. As is common practice in the FDA, QT/QTc data are analyzed and interpreted concurrently by pharmacometricians, clinicians, and statisticians. In general, sponsors conducting such studies will wish to confirm, beyond that observed for placebo after correcting for baseline, (a) the positive control prolongs QTc to a mean effect consistent with its labeling, and (b) the low (therapeutic) and high (supertherapeutic) doses of study drug do not prolong mean QTc per the ICH criteria.

The positive control should have moderate QTc prolongation effect and be less drug concentration dependent. Moxifloxacin, a drug with mild QTc-prolonging properties, is commonly utilized as a positive control in the design of definitive QTc trials and maximum mean changes will be of primary regulatory interest. Therefore, positive control should have average effect sizes of approximately 5–10 ms. The wish is that the positive control will always yield a measured mean effect equivalent to its labeling. However, this wish is not appropriate when one considers measurement and sampling errors. Kenna et al. (13) reported that a recent review of nine QT trials with Moxifloxacin as positive control demonstrated a large variation even for

this known drug; the $\Delta\Delta$QTc ranged from −0.71 to 14.85 ms. Two studies conducted by the same sponsor on Moxifloxacin were not consistent. This does not seem to be due to sample size. Positive control in studies even involving more than 60 subjects should still not be expected to yield point estimates uniformly consistent with QTc labeling for the positive control agents. Mean effect sizes should be expected to vary largely due to random variation and precision of measurement. Hence, large deviation from positive control label could be expected. This complicates interpretation, but also reemphasizes the need of a positive control.

Another design consideration when including a positive control is to try to reduce the frequency when the positive control fails to demonstrate a QTc effect of the expected magnitude with the expected precision. A failure to demonstrate the anticipated effect of the positive control can cast doubt on the findings related to the investigational drug, especially if the study suggests no QT/QTc prolongation for the latter. On the other hand, if the study shows a significantly larger increase in QT/QTc interval among subjects receiving the investigational drug than those receiving the placebo, this finding is likely to stand and be requested to be labeled along with the positive control data.

One argument in support of the positive control requirement is that since it is not reasonable to assume that all factors will remain constant from study to study, a positive control is the best guard against all known and unknown factors. On the other hand, if a sponsor has employed the same process and used the same facilities to collect ECG recordings for the TQT study across many development programs, the sponsor might conceivably be able to conduct some validation studies periodically to reaffirm the process, thus avoiding the need to include a positive control in every definitive QTc study. While this line of thinking has merits, there has not been much experience with its implementation.

For $\Delta\Delta$QTc, presumably one wishes to rule out that an increase of 5 will result when the study drug is administered; hence, attention is on the upper bound of the confidence interval. However, as discussed earlier, mean effect sizes for study drug baseline-corrected comparisons to placebo should also be expected to vary between −5 and 10 ms due to random variation and precision of measurement. Hence, a false-positive result could also be observed simply due to random chance.

In summary, interpretation of mean effect data arising from such studies is not trivial, and the potential for a variety of false-negative and false-positive findings is a reality. Given significant current uncertainty, further regulatory, clinical, and pharmacometric research should be conducted to deal with these issues. The FDA is actively working with sponsors, the pharmaceutical industry, and academia to best handle these issues.

40.3.6 Analysis of Outliers

It is commonly stated that the analysis of extreme values often plays a more important role than that of the average values in clinical trials because it provides more information on the extent of safety concern at the individual level (14). Extreme values can be examined by creating frequency distributions for maximum absolute values as well as maximum increases from baseline (correcting for placebo), using reference limits of 450 and 500 ms on QT or 30 and 60 ms on $\Delta\Delta$QTc. It is important to account for covariates known to affect the distribution of QT/QTc values

in the population (e.g., age and gender). Outlier analysis is significantly affected by the HR correction method, number of replications at each time point, and total number of time points in the collection profile. When there are replicated ECGs at each time point for both baseline and treatment arm, the calculation of outliers is complicated by the three methods of calculation: determining the outlier from the difference of means of both treatment and baseline, subtraction of the mean of baseline measurements from the individual observation of treatment arm, and from all pairs of data subtractions. Certainly, the last situation will yield the highest rate of outliers. There is no common agreed on method for this calculation yet. The common practice is to calculate the means at both treatment and baseline first, then determine the outlier from the difference between the means.

One possible disadvantage of the categorical analysis above is that it can mask repeated occurrences of pathological QT/QTc measurements or changes within a subject, so that similar supplemental tables can be constructed for frequency distributions of the number of ECGs rather than number of subjects. It is important that as the number of ECG recordings per subject increases, one is more likely to observe large values of the QT/QTc interval or large changes from baseline. Therefore, a control group with the same number of ECGs per subject is required to properly interpret the results of these categorical analyses. Example presentations of outliers are given in Tables 40.3 and 40.4.

TABLE 40.3 Presenting Outliers: The Change of QTcF > 450 ms

Regimen	N(A)	N(B)	N Outliers TRT A (%)	N Outliers TRT B
M	3537	1751	19 (0.54)	31 (1.77)
M3	—	782	—	24 (3.07)
M5	—	750	—	21 (2.80)
P	3423	1667	5 (0.15)	1 (0.06)
P3	—	1563	—	10 (0.64)
P5	—	751	—	3 (0.40)
S10	4957	—	68 (1.37)	—
S30	3060	—	181 (5.92)	—

TABLE 40.4 Presenting Outliners: The Change of QTcF from Baseline >30 ms and >60 ms

Regimen	N(A)	N(B)	>30 and <60 ms		>60 ms	
			TRT A	TRT B	TRT A	TRT B
M	638	289	23 (3.6)	18 (5.9)	0 (0.0)	0 (0.0)
M3	—	232	—	53 (18.5)	—	1 (0.3)
M5	—	232	—	43 (15.6)	—	0 (0.0)
P	625	304	2 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)
P3	—	262	—	24 (8.4)	—	0 (0.0)
P5	—	267	—	43 (15.6)	—	0 (0.0)
S10	1060	—	128 (10.8)	—	0 (0.0)	—
S30	880	—	239 (21.3)	—	3 (0.3)	—

40.3.7   Automation of Data Analysis

As many new drugs are subjected to a TQT study, the FDA will receive more and more information from sponsors. In order to capture all the data in submissions in a single location for future reference, and for future data analysis, a database with an analytical tool—the Qtech—was developed in the Office of Clinical Pharmacology and Biopharmaceuticals within the FDA (15). The QTech is a Visual Basic

FIGURE 40.7   The QTech platform for QT data analysis automation.

Application that works in combination with S-Plus, MS Access, MS Excel, and MS Word to provide the users with many functionalities such as generating powerful graphs and reports in addition to storing, loading, querying, and analyzing data in the database. (Minimum Software requirements are S-Plus 6.0, Access 2002 with Service Pack 3 (SP), Excel 2002 with SP3, and Word 2002 with SP3). The platform is presented in Figure 40.7.

There are two important features in QTech: (a) loading data into the database and (b) data querying and analysis. By selecting one of the tasks, the user can go directly to the task at hand. New improvement of the software is to include a drop-down menu for more functions and a link to S-Plus for fast loading of data and analysis. Industry and academia are encouraged to use this FDA in-house software as a template to develop other types of automation for the QT assessment world.

40.4   SUMMARY

This chapter focuses on pharmacometric methods used in QT data analysis. Pharmacometric issues in the design and analysis of thorough QT clinical trials (TQT), and the methods used to assess QT prolongation potential are described. Analysis methods discussed are patterned after the new International Conference on Harmonisation (ICH) guidance for clinical QTc assessment and are illustrated with data from real clinical trials. Issues in the interpretation of data are also discussed.

REFERENCES

1. S. G. Prior, Risk stratification in long QT syndrome. *N Engl J Med* **348**:1866–1874 (2003).
2. QT/QTc Prolongation: Report from the PhRMA Statistics Expert Team, 14 August 2003, Investigating drug-induced QT and QTc prolongation in the clinic: statistical design and analysis considerations. *Drug Inf J* **39**:243–266 (2003).
3. H. C. Bazett, An analysis of time relations of electrocardiograms. *Heart* **7**:353–367 (1920).
4. L. S. Fridericia, Die Systolendauer im Elecktrokardiogramm bei normalen Menschen und bei Herzkranken. *Ada Med Scand* **53**:469–486 (1920).
5. M. Malik, Problems of heart rate correction in the assessment of drug-induced QT interval prolongation. *J Cardiovasc Electrophysiol* **12**:411–420 (2001).
6. C. Funck-Brentano and P. Jaillon, Rate-corrected QT interval: techniques and limitations. *Am J Cardiol* **72**:17B–22B (1993).
7. International Conference on Harmonisation (ICH) E14, *Guidance for Industry*, *E9 Statistical Principles for Clinical Trials*. `www.fda.gov/cder/guidance/` (1998).
8. M. Malik and A. J. Camm, Evaluation of drug-induced QT interval prolongation. *Drug Saf* **24**:323–351 (2001).
9. H. Sun, P. Chen, L. Kenna, and P. Lee, The chaotic QT interval variabilities on risk assessment trial designs. *Clin Pharmacol Ther* **75**(2):P55 (2004).
10. H. Sun, P. Chen, J. Hunt, and H. Malinowski, Frequency of QT recording and the reliability of QT prolongation assessments. *Clin Pharmacol Ther* **75**(2):P48 (2004).

11. M. Agin, D. Kazierad, R. Abel, R. Anziano, W. Billing, G. Layton, J. Mancuso, D. Stricter, W. Xu, R. Blum, and D. Jorkasky, Assessing QT variability in healthy volunteers. Presented at ACCP Annual Meeting, September 2003.

12. S. H. Lee, H. Sun, P. Chen, S. Doddapaneni, J. Hunt, and H. Malinowski, Sensitivity/reliability of the time-matched baseline subtraction method in assessment of QTc interval prolongation. *Clin Pharmacol Ther* **75**(2):P56 (2004).

13. L. A. Kenna, A. Parekh, V. Jarugula, D. J. Chatterjee, H. Sun, M. J. Kim, S. Ortiz, J. P. Hunt, and H. Malinowski, Experience evaluating QT prolongation data. *Clin Pharmacol Ther* **75**(2):P7 (2004)

14. C. Chuang-Stein, Safety analysis in controlled clinical trials. *Drug Inf J* **32**:1363S–1372S (1998).

15. *Quantitative Tools for Electrocardiogram Computation Harmonization* (QTech), FDA, OCPB, Rockville, MD, 2004.

Using Pharmacometrics in the Development of Therapeutic Biological Agents

DIANE R. MOULD

41.1 PHARMACOKINETICS OF THERAPEUTIC PROTEINS

41.1.1 Background

Overall, there are some profound differences between the pharmacokinetic (PK) behavior of biologics and small molecules. Table 41.1 summarizes the major differences between these two broad classes of molecules. When evaluating the PK behavior of any protein, it is important to understand the biology and the pharmacology of the system that the therapeutic biologic is acting on in order to anticipate the expected covariates and behavior of the drug.

Initially, the structure of the protein that is being developed can provide some information as to the likely clearance and also can suggest feasible routes of administration. Table 41.2 presents the relationships between the molecular weight (MW), bioavailability, and clearance of biologics.

In the development of any therapeutic biological agent, the ability to anticipate the PK and pharmacodynamic (PD) behavior of the agent is helpful. However, the PK behavior of proteins is quite distinct from the behavior of small molecules and the drug development path for biologics is not as standardized as it is for small molecules. Table 41.3 gives a general overview of some characteristics of proteins that can help the analyst anticipate the likely PK behavior of a novel biological agent before it goes into clinical testing.

41.1.2 Protein Structure

The architect Louis Sullivan first coined the phrase "form follows function." In the biological setting, however, the reverse is true: function follows form. During any PK or PD assessment of therapeutic proteins, the size and structure of the molecule

TABLE 41.1   Overview of Pharmacokinetic Differences Between Chemical Entities and Proteins

Chemical Entities	Therapeutic Biologics
Pharmacokinetics usually independent of pharmacodynamics	Pharmacokinetics often dependent on pharmacodynamics
Usually linear pharmacokinetics	Often nonlinear pharmacokinetics
Metabolic breakdown	Proteolytic breakdown
Renal clearance often important	Renal clearance uncommon if MW higher than 50 kD
Free concentrations useful ("coverage")	Free concentrations *may* cause problems (immunogenicity)
Binding implies distribution	Binding implies clearance
Tissue penetration often good	Usually poor tissue penetration
PK drug interactions possible	PK drug interactions rare
PD drug interactions rare	PD drug interactions possible

TABLE 41.2   Relationships Between Molecular Weight, Bioavailabilitys, and Clearance

MW < 20 kD	20 kD < MW < 50 kD	MW > 50 kD
Good bioavailability when given SC	Adequate bioavailability when given SC	Poor bioavailability when given SC
Can be given via inhalation	Inhalation possible	Not suitable for inhalation
Usually very fast clearance (minutes to hours)	Usually moderate clearance (hours)	Moderate to slow clearance (hours to days)

needs to be taken into account in order to develop appropriate models for both the pharmacokinetics and pharmacodynamics of a biological agent.

Proteins are characterized by their primary, secondary, tertiary, and quaternary structures. The primary structure is the sequence of the amino acids in the polypeptide chain that makes up the protein. Secondary structure refers to the first folding of the amino acid chain and reflects, for example, disulfide bonds. Tertiary structure (a monomer) is the final folded configuration of the protein that is controlled by the primary and secondary structures and is thermodynamically driven by the relative hydrophobicity of the component amino acids in the structure. Quaternary structure refers to the functional association of several polypeptides (monomers). For example, the final structure of hemoglobin consists of four associated monomers. Any change in the primary structure of a protein often results in changes to all the higher level structure. Protein structures must be characterized and controlled during the production process.

To give some specific examples of how physical structure plays a role in the pharmacokinetics of a therapeutic protein, first consider a monoclonal antibody binding fragment (Fab), which is the active binding region of an antibody, and Fab fragments, which have been extensively investigated as potential therapeutic agents but because they are rapidly cleared, their potential as therapeutic agents is limited. Reengineering the structure by replacing a single disulfide bridge between Fab arms with a thioether bridge increased the mean residence time of the fragment

TABLE 41.3 Lookup Table of Protein Characteristics and Associated PK Behavior

Question	Answer											
Human structure	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Receptor-mediated clearance	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Changes receptor number	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Functional Fc present	No	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Clearance	Linear, fast, constant N-NAB	Linear, fast, constant NAB	Linear, slow, constant N-NAB	Linear, slow, constant NAB	Nonlinear, fast, constant N-NAB	Nonlinear, fast, constant NAB	Nonlinear, fast, variable N-NAB	Nonlinear, fast, variable NAB	Nonlinear, slow, constant N-NAB	Nonlinear, slow, constant NAB	Nonlinear, slow, variable N-NAB	Nonlinear, slow, variable NAB
Consider modification[a]	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Example	Human enzymes, Epogen, factor VIII, Fuzeon	Porcine insulin vaccines	Protein construct, "fusion protein"		Glucocerebrosidase, Ceridase, and Cerizyme	Fab fragment, single-chain antibody	GCSF, TPO	OKT3	"Coating antibody," Enberel, Clenoliximab	Chimeric coating antibody	Herceptin, Campath	Chimeric lytic antibody

[a]Refers to modifications such as PEGylation or hyperglycosylation, which alters pharmacological behavior.

in normal mice by threefold (1). A second example of a structural change resulting in altered pharmacokinetics is Tenecteplase, a fibrinolytic protein developed from human tissue plasminogen activator (alteplase) for the treatment of acute myocardial infarction. Specific mutations at three sites in the original alteplase molecule resulted in 15-fold higher fibrin specificity, 80-fold reduced binding affinity to the physiological plasminogen activator inhibitor PAI-1, and sixfold increase in the half-life (2). Clearly, the structure of the protein is important to both the pharmacokinetics and pharmacodynamics of any biological therapeutic agent, emphasizing the need to characterize the protein structure at all levels and control it adequately during manufacturing. Some of the basic aspects of protein production and engineered structural changes are presented in the following sections.

41.1.2.1   *Production of Therapeutic Proteins*
Bacteria and Chinese hamster ovary (CHO) cell lines are commonly employed for the production of recombinant therapeutic proteins. The process given below is used in the development and production of therapeutic antibodies, which constitute a majority of the proteins being used clinically today.

In 1975, Kohler and Milstein (3) presented a method for preparing murine cell cultures that would produce antibodies targeted against a specific antigen. Producing mouse antibodies to selected antigens is very easy to do, and murine antibodies have been shown to have clinical utility, although all murine antibodies have been associated with the formation of human anti-murine antibodies (HAMA).

This method ultimately led to the development of Orthoclone muromonab-CD3 (OKT3®), the first monocloncal antibody approved for use in humans. OKT3 was a murine IgG2 antibody that binds and modulates the CD3 receptor site on cytotoxic T-lymphocytes, interfering with antigen recognition and preventing cellular proliferation. OKT3 is used for the treatment of acute rejection in renal transplantation. However, it has since been determined that antibodies that have murine structure are more immunogenic than ones that have been engineered to have human structure. This is because foreign proteins are recognized as such and consequently elicit an antibody response against them (e.g., HAMA) or more broadly human anti-globulin antibody (HAGA). In general, HAMA response is polyclonal, with increased levels of IgM and IgG that are directed against the mouse-specific determinant, the isotype (the heavy chain or FC portion), the binding region (F(ab')2), and the "idiotype" of mouse immunoglobins (4). The development of neutralizing antibodies directed against the foreign protein restricts their usefulness for several reasons: the development of HAMA can result in anaphylaxis or other related adverse events (e.g., fever, chills, serum sickness, anemia, leukopenia, arthralgia, rash), and HAMA forms a new, and fast, route of elimination that removes the therapeutic antibody from circulation, which reduces the circulating therapeutic protein concentrations and consequently its beneficial effect (5). Because of the HAMA response, murine antibodies are not suitable for treatment durations exceeding 10 days and cannot be administered at some later time to a patient who was previously exposed because of the potential for anaphylaxis.

41.1.2.2   *Humanization*
Structural changes in proteins can result in substantially different PK behavior. Protein engineering using recombinant DNA technology has provided a partial

solution to the problems associated with HAGA by developing methods to construct chimeric genes that fuse rodent exons for the monoclonal antibody variable regions with human exons encoding the heavy and light chain constant (Fc) regions of the antibodies (6, 7). These constructs produce antibodies with human effector functions due to the human Fc portion and, at the same time, theoretically reduce the likelihood of HAGA response to a major part of the protein. The transfer of murine binding regions into human frameworks transfers the ability to recognize the antigen but provides a slightly less immunogenic framework. In addition to reducing immunogenicity, the use of the human constant (Fc) region is theoretically associated with improved effector function of the therapeutic antibody (e.g., complement protein and antibody-dependent cell cytotoxicity (ADCC)) as well as decreased clearance due to the improved ability of the antibody to take advantage of the body's tendency to conserve antibodies through protective mechanisms such as Brambell receptors. In summary, the more "human" the antibody structure, the less immunogenic the agent is, the longer the half-life, and the greater the likelihood of utilizing ADCC.

The benefits of humanization are not completely straightforward. Many features of immunoglobulin sequences are conserved between species and thus there is no concept of an immunoglobulin sequence appearing to be completely murine. Therefore, the humanization of antibodies does not automatically preclude the development of HAGA (8) because the variable regions of the antibody are still murine and therefore chimeric and humanized proteins may still develop HAGA. In addition, there is considerable homology between many murine and human variable region sequences, making the development of HAGA difficult to predict.

41.1.2.3 *PEGylation*
One of the drawbacks of biological agents is the need for parenteral routes of administration. In some cases, these agents are administered as subcutaneous injections, which reduces the number of clinic visits required for treatment. However, many of these agents require frequent dosing because of their short half-life. The discomfort associated with frequent injections can negatively impact patient compliance and there are other issues associated with the disposal of used syringes.

The concept of modifying therapeutic molecules through the covalent attachment of poly(ethylene glycol) (PEG) moieties (PEGylation) was first introduced by Abuchowski et al. (9) in 1977. This approach has been shown to be effective in decreasing the clearance of therapeutic protein agents, as well as reducing the incidence of neutralizing antibody formation, the mechanism of which is described below.

PEGylation reduces renal and hepatic clearance and, for some products, effectively increasing the circulating half-life of the agent. PEGylation also results in a more sustained absorption after subcutaneous administration as well as restricted distribution. These PK changes usually result in more constant plasma concentrations, which can be maintained near the desired target levels with less frequent dosing.

Additionally, PEG modification may decrease adverse effects caused by the large variations in peak-to-trough plasma drug concentrations associated with intermittent administration and by "covering" the foreign protein (resulting in PEG-induced steric hindrance) can prevent immune recognition and reduce the immunogenicity

as compared to the unmodified protein (10, 11). However, it should be noted that for all covalently PEGylated and successfully marketed PEGylated agents, the weight of the added PEG was at least as great as the weight of the protein being modified.

Modification of a protein by PEGylation also causes changes to the observed PD properties due to altered protein structure and hydrophilicity, which in turn results in different binding properties of the native protein. In general, the binding capacity of PEGylated proteins is reduced as compared to the native protein. Because the size, geometry, and attachment site of the PEG moiety play pivotal roles in observed changes of these properties, therapeutically optimized PEGylated agents must be individually designed.

41.1.2.4 *Hyperglycosylation*

Perhaps the best known example of modification of a protein resulting in altered pharmacokinetics is darbepoetin alfa (Aranesp®) (12). Darbepoetin alfa is a hyperglycosylated analog of recombinant human erythropoietin. The addition of sialic acid residues to erythropoietin resulted in a substantial prolongation of circulating half-life. The terminal half-life of darbepoetin alfa was two to three times longer and the clearance was approximately four times slower than epoetin (13). Similar results have been reported when "glycoengineering" was applied to thrombopoietin and leptin (14). In addition, when L-asparaginase was conjugated with colominic acid (polysialic acid), the immunogenicity was reduced (15). Antibody titers appeared highest for the native enzyme and were generally reduced as the degree of polysialylation increased. In addition, the half-lives of these preparations were three- to fourfold greater than that of the native enzyme.

41.1.3 Clearance

For proteins, structure has an impact on the clearance. For example, the half-life of an intact IgG molecule is 23 days, while for an intact Fc fragment the half-life is 10–20 days (16). Similarly, the binding regions of antibodies (F(ab′)2 fragments) are cleared very rapidly (17). As mentioned previously, the PK behavior can be altered not only by changes in the amino acid sequence but also by changes in the pattern of glycosylation on the protein (18). Consequently, structural changes can alter the PK and the PD behavior of the drug.

Therapeutic proteins can undergo several routes of elimination: renal, hepatic, receptor mediated, and HAGA directed. Not all proteins undergo clearance through all possible routes. Again, the type of elimination is partly dependent on the structure of the protein, its molecular weight, and immunogenicity. In addition, the role that receptor-mediated clearance plays in the overall clearance of a protein depends on the functionality of that protein. An excellent example of characterization of different routes of clearance of a biologic agent is a report of the pharmacokinetics of SB-251353, a low molecular weight protein that is a truncated form of the human CXC chemokine growth-related gene product beta (19). The pharmacology of this agent was studied in the mouse. Primarily, the clearance appears to be mediated by its pharmacologic receptor, CXCR2, through endocytosis with subsequent lysosomal degradation. SB-251353 is eliminated via renal and hepatic routes as well. Microscopic autoradiography showed uptake into renal proximal tubule epithelial

cells with limited excretion of SB-251353 in the urine (<2%). Binding to hepatic sinusoids and connective tissue in the dermis was also observed, which is characteristic of the mechanism of hepatic elimination for proteins.

41.1.3.1 Renal Elimination

The kidney plays an important role in the metabolism of low molecular weight (MW < 20 kD) proteins, which are extensively filtered by the kidney. Proteins that are filtered by the kidney are typically reabsorbed from the luminal side by renal tubular cells and released back to the circulation as intact molecules or as amino acids and peptide fragments. The renal filtration, absorption, and final disposition of three low molecular weight proteins (lysozyme (MW 14 kD), insulin (MW 5.83 kD), and recombinant human growth hormone (MW 20 kD)) were studied in order to gain an understanding of the fundamental variables involved in the renal handling of these agents (20). Maack (20) reported that the glomerular barrier offers little hindrance to the filtration of these low molecular weight proteins. The intrarenal route by which low molecular weight therapeutic proteins accumulate in the kidney is primarily via filtration and uptake by renal tubular cells. Uptake or adsorption of these exogenous proteins from the peritubular side was found to be minor compared to luminal uptake. Current evidence (21) indicates that only the proximal tubule possesses the mechanism for degradation and transport of these proteins and reabsorption of the resulting metabolic products. Proteins that are filtered at the glomerulus are absorbed into apical vacuoles, which fuse with primary lysosomes. The proteins are then hydrolyzed in the vacuoles and the proteolytic products diffuse out of the cell and back into the blood. This process, which conserves amino acids, inactivates toxic substances and regulates circulating levels of protein and peptide hormones. This mechanism plays an important role in the clearance of low molecular weight proteins.

41.1.3.2 Hepatic Clearance

Proteins that are larger than 20 kD undergo relatively little renal elimination as an intact molecule (22) although the mechanisms for renal filtration and breakdown described in Section 41.1.3.1 on renal clearance does play a role for proteins up to 50 kD (23). For proteins that have a molecular weight that is higher than 50 kD, organ-based clearance is primarily limited to clearance by other cellular mechanisms such as through the Kupffer cells and endothelial sinus cells in the liver, which act to remove dead cells and other large proteins from the blood (24–26). Other general routes of clearance of large proteins are via uptake by splenic macrophages and some limited proteolytic clearance in the lung and intestines (27). As with renal clearance, therapeutic proteins are internalized by these cells, are broken down to their component amino acids, and then are salvaged by the cells.

41.1.3.3 Receptor-Mediated Clearance

Many therapeutic proteins, particularly protein hormones and monoclonal antibodies, exhibit saturable clearance mechanisms that appear to be receptor mediated (28, 29). This form of clearance is usually directly linked to the pharmacological activity of the drug and plays an intricate role in removal of the protein from circulation. The mechanism usually involves binding of the therapeutic agent followed either by internalization (pinocytosis) and breakdown of the protein or

alternatively shedding of the receptor–drug complex followed by uptake of the complex by Kupffer cells.

A good example of this is the observation that the clearance of granulocyte colony-stimulating factor (G-CSF) changes over time (30, 31). This study looked at the correlation between G-CSF clearance and absolute neutrophil count (ANC) in patients treated with high-dose chemotherapy followed by autologous bone marrow transplantation after intravenous administration of G-CSF (5 or 16 μg/kg/day) on three separate days. G-CSF plasma clearance increased with time post-transplant. Regression analysis of G-CSF clearance with neutrophil count revealed a strong and statistically significant linear relationship. The role of neutrophils in G-CSF clearance was also evaluated in vitro using polymorphonuclear neutrophils (PMNs) incubated with G-CSF. At low G-CSF concentrations in vitro, there was an increase in G-CSF clearance with increasing PMNs, but at higher G-CSF concentrations this relationship did not hold. The correlation between G-CSF clearance and ANCs both in vivo and in vitro is consistent with the hypothesis that receptor-mediated clearance by neutrophils is one of the major pathways of G-CSF clearance.

Another example of receptor-mediated clearance is the glycoprotein IIb/IIIa antagonist abciximab, which is administered to patients undergoing percutaneous transluminal coronary angioplasty (32). Concentration–time profiles of this drug showed rapidly decreasing plasma abciximab concentrations at early times postdose, but the terminal disposition phase was prolonged, suggesting saturable behavior. The PK model included both receptor binding and linear clearance mechanisms.

41.1.3.4 *Clearance by Neutralizing and Nonneutralizing Antibodies*

During treatment with a therapeutic protein, antibodies can be formed against the therapeutic agent. In fact, this is the case with nearly all biopharmaceuticals, even small molecules. The frequency and consequence of the formation of antibodies directed against the therapeutic protein varies widely. Antibody formation against a therapeutic protein has profound consequences on the pharmacokinetics and potentially the pharmacodynamics of the biological agent. It is important to understand not only that these antibodies form, but also the factors relating to their formation. There are two forms of antibodies that can appear during administration of a therapeutic biologic agent: nonneutralizing antibodies (N-NAbs) and neutralizing antibodies (NAbs). The definitions for neutralizing and nonneutralizing antibodies are usually set up for each new biologic under development and these two antibody types are often distinguished on the basis of in vitro assays. Typically, neutralizing antibodies are defined as antibodies that block the activity of the therapeutic protein in an in vitro biological assay of activity. However, it should be noted that a lack of neutralizing activity in vitro does not always correspond perfectly with neutralizing activity in vivo.

Both nonneutralizing and neutralizing antibodies can form complexes with the therapeutic protein, which are subsequently taken up by leukocytes in the peripheral blood, resulting in loss of active drug and its associated clinical activity. There are numerous mechanisms by which these antibodies work. For example, neutralization can occur by inhibiting the attachment of the therapeutic agent to its target receptor when the neutralizing antibody binds directly at, or near, the target site.

The binding of the neutralizing antibody can either physically block the site or cause a conformational change so that the therapeutic agent and the receptor can no longer interact. Most often, however, the neutralization is thought to be mediated by the direct binding of the neutralizing antibody to the therapeutic agent. Neutralization can also occur by causing aggregation of the therapeutic agents. This activity can be seen with any kind of antibody directed against the therapeutic agent, whether it is neutralizing or not. Another site of neutralization can occur when the neutralizing antibody binds the therapeutic agent following attachment of that agent. This can occur due to the formation of a new epitope, which is either exposed or formed following the binding between the therapeutic agent and the target receptor. There are numerous other proposed mechanisms of interaction between antibodies and therapeutic protein.

The development of neutralizing antibodies appears to be dependent on both the structure of the protein (e.g., how "foreign" the protein appears to the body) and the route of administration (33), with intravenous and local routes usually having a lower incidence of antibodies than subcutaneous or intramuscular routes of administration. Inhalation appears to be the most immunogenic route of administration. In addition, the dose regimen used is also important with infrequent or intermittent dosing generally being more immunogenic than more frequent dosing although the pattern is less clear. Much of the information on development of neutralizing and nonneutralizing antibodies comes from research on vaccines, but appears to be broadly applicable to other therapeutic proteins as well.

Essentially, there are two mechanisms behind the formation of antibodies directed against an exogenously administered protein. The first mechanism is the standard reaction of a human body to foreign protein, such as those arising from animal, bacterial, or plant origins (e.g., OKT3). This reaction is comparable with the administration of a vaccine. Neutralizing antibodies against the foreign protein appear rapidly in the majority of cases, sometimes even after a single injection, and can persist for a long time.

The other mechanism by which antibodies are induced is based on breaking the immune tolerance that normally exists to self-antigens. This is the mechanism leading to the antibodies directed against recombinant human homologues such as interferons or thrombopoietin. These antibodies are usually nonneutralizing and generally, albeit rarely, appear after prolonged treatment.

The structure of the protein, including the glycosylation pattern, as well as the presence of foreign (e.g., murine) epitopes can cause an immunogenic response. In the case of glycosylation changes, the protein may be less soluble than the endogenous form, or epitopes that are normally hidden by glycosylation become exposed, rendering the product immunogenic (34).

Patient characteristics can also be an important factor in determining whether antibodies will be produced. For example, cancer patients, whose immune response may be impaired due to their disease and treatment with myelosuppressive chemotherapy, would presumably have a lower incidence of antibodies, as might any patient whose disease or treatment lowers immune response. It should be noted, however, that even cancer patients are capable of expressing antibodies to foreign proteins (35).

The effect of patient gender can also be important for immunogenicity. A report of the emergence of alpha-Gal A antibodies in patients with Fabry disease (36)

found that during the first 6–12 months of intravenous administration of recombinant enzymes (rh-alpha-Gal A), the female patients did not develop detectable amounts of antibodies. Conversely, after 6 months, 11/16 male patients showed high titers of immunoglobulin G antibodies that cross-reacted with both recombinant enzymes. These antibodies were able to almost completely neutralize in vitro activity. All antibody-negative patients showed a significant improvement as compared to antibody-positive patients whose condition either remained stable or worsened.

In addition to increased clearance, the formation of neutralizing antibodies has been shown to result in severe and prolonged adverse events (37, 38). This has been found for both Eprex® (a formulation of erythropoietin used to improve red cell count in the anemia of renal failure) and for recombinant human thrombopoietin (a cytokine involved in the development and maturation of platelets). In one report (39), serum samples from 13 patients who had developed pure red-cell aplasia while treated with Eprex were tested for neutralizing antibodies that could inhibit erythroid-colony formation in vitro. The presence of anti-erythropoietin antibodies was detected in all patients evaluated. In all of the patients, the antibody titer slowly decreased after the discontinuation of treatment with erythropoietin and the anemia resolved. Similarly, subjects who had been administered recombinant thrombopoietin developed neutralizing antibodies against endogenous thrombopoietin, which resulted in thrombocytopenia. The development of neutralizing antibodies against endogenous proteins is of concern during the development of any biological agent.

41.1.3.5   *Brambell Receptors and Antibody Recycling*

For the most part, Fc receptors are associated with activating host cell defense mechanisms, initiating cell signaling when an antibody binds to a target cell-surface antigen. However, there is a second class of Fc receptor that is involved in antibody catabolism. The receptor, called FcRn for the neonatal Fc receptor, is also sometimes referred to as the Brambell receptor. The existence of this receptor was first proposed by Brambell (40) in 1966 to describe the mechanism by which maternal IgG was transferred across the neonatal gut. These receptors are also involved in maintaining homeostasis of circulating IgG in the body (41–42). FcRn is expressed in endothelial cells and removes circulating IgG from the blood by pinocytosis. The IgG that is removed by this process can bind to the FcRn receptors present in the vesicle. The IgG–FcRn complex is protected from degradation by lysosomal enzymes and recycled back to the surface where the IgG is released back into the blood. A schematic of this process is presented in Figure 41.1.

Human IgG1, IgG2, and IgG4 all have high affinities for human FcRn, which is why these molecules have a long half-life. IgG3 has a lower affinity for FcRn and consequently exhibits a shorter half-life. The alteration of specific amino acids or glycosylation patterns in the constant (Fc) domain can therefore result in altered half-lives for immunoglobins by altering their binding affinity for FcRn (43). In addition, FcRn can become saturated at high concentrations of antibody, resulting in a loss of the protective mechanism and a subsequent increase in the clearance or shortening of the half-life. The impact of this on the overall clearance of a monoclonal antibody is shown in Figure 41.2.

FIGURE 41.1   Schematic diagram of the role of brambell receptors in the salvage of antibodies. **A**—Cells with FcRn and two antibodies, the light colored antibody has a functional Fc receptor and the dark one does not. **B**—The antibodies bind to the FcRn receptors. **C**—The bound antibodies are internalized by pinocytosis. **D**—The environment of the internalized antibodies changes, causing the dark antibody (with a nonfunctional Fc receptor) to disassociate from the FcRn receptor, whereas the antibody with the functional receptor remains bound. **E**—The bound antibody is protected from proteolytic degradation, whereas the disassociated antibody is broken down. **F**—The FcRn receptors cycle back to the cell surface. **G**—The cell is freed by releasing the intact functional antibody back to circulation.

41.1.4   Absorption

41.1.4.1   Overview of Absorption

At present, most therapeutic biologics are administered parenterally, either via IV infusion or as a subcutaneous (SC) injection. Other routes such as intramuscular injection and inhalation are used but are much less commonly seen in practice. Not much is known about the mechanism of absorption of proteins administered via SC or intramuscular injection, although the majority of uptake is presumably primarily via the lymphatic system (44, 45). Supersaxo et al. (44) investigated the lymphatic absorption of four compounds with different molecular weights administered subcutaneously in sheep. Lymphatic uptake was determined by measuring drug recovery in lymph. The cumulative recovered percentage of administered dose showed that in the investigated MW range, there was a positive linear relationship between the molecular weight and the proportion of the dose absorbed via the lymphatic system. Molecules with MW > 16 kD were primarily absorbed via the lymphatic system.

An interesting aspect of this mechanism of absorption is the link between the pharmacokinetics and PD activity of some drugs. A study of the pharmacokinetics and pharmacodynamics of G-CSF (46) found no correlation between C_{max} and

FIGURE 41.2 Clearance of antibodies is dependent on concentration. At low concentrations, therapeutic proteins may exhibit concentration-dependent clearance, with the clearance decreasing as concentrations increase. For most biological agents that display this behavior, the clearance will decrease to a particular level (the nadir in the curve) and then remain constant. However, for antibodies, a second form of saturation can occur, where the Brambell receptors become saturated, resulting in an apparent increase in clearance as concentrations increase.

increases in neutrophil count, but there was a negative correlation between AUC and neutrophils. The measured G-CSF concentrations reflect the fraction of drug that escapes clearance in the lymphatic system prior to entering the circulation (somewhat analogous to first pass metabolism), and with continued dosing the relative bioavailability of G-CSF would be expected to decrease.

41.1.4.2 *Relationship Between Molecular Weight and Bioavailable Fraction*

When administered via SC, intramuscular, or inhalation routes, the bioavailability of therapeutic proteins is variable and the fraction absorbed is dependent on the molecular weight of the protein (47). Interferon alpha, which is a relatively low molecular weight protein (19 kD), has good bioavailability following SC administration (80%), whereas most therapeutic monoclonal antibodies have bioavailability of approximately 20–60% following SC administration.

The reason that proteins generally exhibit low and variable bioavailability is not presently understood, although several mechanisms such as degradation, aggregation, or metabolism at the site of injection have all been proposed (48, 49); however, there is little evidence supporting these theories.

41.1.5 **Volume of Distribution**

For large proteins, transfer across cell membranes is limited due to size and hydrophilicity of the molecule. There is some evidence that Brambell receptors may play a

TABLE 41.4   Pharmacokinetic Parameters for Several Marketed Therapeutic Monoclonal Antibodies and Derivatives

Generic Name	Brand Name	Type	Target Antigen	V_{ss} (L)
Abciximab	Repro	Fab fragment	GP IIb/IIIa	8
Basiliximab	Simulect	Chimeric IgG1	CD25	9
Bevacizumab	Avastin	Humanized IgG1	VEGF	—
Cetuximab	Erbitux	Humanized IgG1	EGF receptor	4.4
Daclizumab	Zenapax	Humanized IgG1	CD25	6
Etanercept	Enberel	Fusion protein linked to the Fc portion of human IgG1	TNF	—
Gemtuzumab	Myelotarg	Humanized IgG4	CD33	20
Infliximab	Remicade	Chimeric IgG1	TNF	3
Trastuzumab	Herceptin	Humanized IgG1	Her2-neu	4

role in facilitating crossing cell membranes for macromolecules that have functional Fcγ binding (40), but the extent of this activity appears to be limited. Karanikas et al. (50) demonstrated that there is little cellular penetration of monoclonal antibodies, even in cells carrying target receptors. Lin et al. (51) evaluated the distribution of a recombinant humanized IgG1 monoclonal antibody (MAb) directed against vascular endothelia growth factor (VEGF) in rabbits (51). These findings showed that, as expected, serum concentrations of the MAb were 10 times higher than the highest tissue concentration. Furthermore, after 24 hours, evaluable autoradiography was limited due to the recycling of the labeled amino acids by the body, making assessments of tissue distribution difficult.

Consequently, most high molecular weight therapeutic proteins (MW > 50 kD) appear to have a distributional volume on the order of 0.1 L/kg, which is approximately equal to the extracellular fluid volume. The volumes of distribution of several marketed therapeutic monoclonal antibodies are provided in Table 41.4. The values for these parameters were taken from the *Physician's Desk Reference* (52), which includes labeling information for each agent. Lower molecular weight proteins (MW > 50 kD) generally have a slightly higher volume of distribution, with the volumes ranging from 0.2 to 0.8 L/kg.

41.2   EVALUATING PHARMACOKINETICS USING MODEL-BASED ANALYSIS

41.2.1   Pharmacokinetic Models

In general, the pharmacokinetics of most therapeutic proteins can be described using either a one- or a two-compartment model. This behavior is dependent in part on the route of administration and the molecular weight of the protein. In general, most proteins display nonlinear clearance or there are parallel linear and nonlinear routes of clearance. Absorption following SC administration is often not straightforward to describe, with many analysts using parallel routes of uptake or modeling the absorption as a slow first-order process.

41.2.1.1 *Absorption Models*

Following SC administration, absorption is generally variable and is often described using parallel routes of uptake (53, 54), although simple first-order models have been used successfully (55). However, the absorption of a biological agent usually follows a complex process. Several simple alternative functions are provided below that have been used to describe the absorption kinetics of various biologics. There are many other functions that can be tested during the model development process.

Inverse Gaussian Input Function   The inverse Gaussian input function is described as follows:

$$Input(T) = Dose \cdot F \cdot \sqrt{\frac{MAT}{2\pi \cdot NV^2 \cdot T^3}} \exp\left(-\frac{(T-MAT)^2}{2 \cdot NV^2 \cdot MAT \cdot T}\right) \quad (41.1)$$

In this equation, *Dose* is the administered dose, *F* is fraction absorbed, *MAT* is the mean input time or mean absorption time, NV^2 is the normalized variance of the Gaussian density function, and *T* is the modulus time following administration of a dose.

Spline Input Function   A cubic spline function can be used to reproduce the input function described by the inverse Gaussian function. This function is somewhat simpler to code than the inverse Gaussian but is also an empirical function. The spline function is

$$Input = \lfloor A \cdot T^3 - B \cdot T^2 + C \cdot T \rfloor \quad (41.2)$$

In this equation, *T* is the relative time postdose, and *A*, *B*, and *C* are the coefficients of the cubic spline. Unlike the inverse Gaussian function, the spline input function does not rely explicitly on dose and can be evaluated both as an explicit input function and as an infusion rate.

Biexponential Input Function   A biexponential (Bateman) function is sometimes useful because it is also capable of producing an input profile similar to the inverse Gaussian function and can describe nonlinear absorption processes. The representation for this input function is given as

$$Input = K_{a0} \cdot \{\exp[-A \cdot (T - Alag)] - \exp[-B \cdot (T - Alag)]\} \quad (41.3)$$

In this equation, K_{a0} is the basic absorption rate, *T* is the modulus time postdose, *A* and *B* are coefficients of the exponential input curves, and *Alag* is the lag time before absorption begins. The input is 0 at all modulus times less than *Alag*.

Surface Input Function   One additional function can be considered when investigating absorption models that might describe a nonstandard absorption process. This is a simple surface function. This input function is represented as follows:

$$Input = K_{a0} \cdot [1 - Alpha \cdot (T - Alag)] \quad (41.4)$$

In this equation, K_{a0} is the basic absorption rate, T is again the modulus time post-dose, *Alag* is the lag time prior to the onset of absorption, and *Alpha* is a coefficient defining the input function curvature.

41.2.1.2   Clearance Models

As might be expected, many biological agents exhibit nonlinear behavior. In some cases, the clearance can be described using parallel linear and nonlinear mechanisms of clearance (19, 53, 56). In order to be able to develop this model, data must be available from a wide range of doses. However, depending on the mechanism of action, the mechanism of clearance, and other aspects such as patient covariates, clearance for some proteins has been described using wholly linear or nonlinear mechanisms. In addition, time- or receptor-dependent clearance mechanisms have also been utilized to explain time-dependent changes in clearance due to the effect of repeated administration of therapeutic proteins. The most common form of clearance is represented as

$$Clearance_{\text{Total}} = \left(\frac{V_{\max} \cdot Concentration}{K_m + Concentration} \right) + CL_{\text{Linear}} \tag{41.5}$$

In this equation, $Clearance_{\text{Total}}$ is the sum of nonlinear and linear clearance. $V_{\max}$ is the maximum nonlinear clearance, K_m is the concentration required to achieve half-maximal nonlinear clearance, and *Concentration* is the drug concentration. CL_{Linear} is the linear component of clearance.

41.2.2   Potential Covariates

During the development of many therapeutic proteins, many of the standard Phase 1 pharmacology studies such as those conducted in special populations (e.g., elderly subjects, renally impaired subjects) and drug interaction studies can be omitted. This is done for numerous reasons, including the fact that administration of protein-based agents can cause the formation of antibodies, which could potentially affect that subject's treatment should they later develop a disease requiring therapy using a similar biologic. Furthermore, in many cases, normal volunteers do not have high levels of receptors for these agents because they do not have the disease, so the basic PK behavior of a biologic agent cannot always be translated between a normal volunteer and a patient. Finally, patients are a more heterogeneous group than normal volunteers, making accrual of appropriate patients for such studies difficult, and if the drug has shown efficacy it may be inappropriate to conduct single-dose studies in these patients. These factors limit the number of studies that are conducted during the development of a biological agent and place a greater importance on the use of population PK modeling to assess the effects of covariates on the PK behavior of these agents. A discussion of the potential covariates that should be considered during a population-based evaluation is given in the following sections.

41.2.2.1   Number of Receptors

For therapeutic proteins that are cleared by receptor-mediated binding, the number of receptors is usually one of the major covariates. Receptor density or receptor-positive cell count has been identified numerous times as a covariate for cytokines

and peptide hormones as well as antibodies (49). In some cases, information on receptor density or number of receptor-positive cells is not available. In these cases, treatment duration can sometimes be used to account for changing (usually decreasing) receptor density over time with the consequent decrease in clearance as treatment with the biologic progresses. An example of this is a population PK evaluation of alemtuzumab (Campath®) (57), which found that the concentration–time profile was best described by a two-compartment model with nonlinear elimination. Campath is cleared by binding to $CD52^+$ cells, which was a strong covariate on the maximum velocity for clearance (V_{max}). During treatment, the number of $CD52^+$ cells is markedly decreased, resulting in decreased clearance of this agent.

41.2.2.2   Patient Characteristics

Patient characteristics can have a profound effect on the pharmacokinetics of therapeutic proteins. For patients undergoing immunosuppressive therapy or those who have a disease that compromises their immune response, antibody formation against the biologic agent is often blunted, delayed, or of lower frequency. However, such patients can and do form antibodies against biologics. As mentioned previously, cancer patients who are treated with asparaginase can develop antibodies against this protein, although PEGylated asparaginase appears to ameliorate the immune response, prolonging the duration of effective treatment with this agent (35).

For proteins that undergo receptor-mediated clearance, the stage of the disease can be a predictive covariate when receptor number is missing. Presumably, patients with more advanced disease would have a tendency to have a greater number of receptor-positive cells. In a categorical sense, patient disease state or disease status may be useful as an explanatory covariate when receptor number is not available. The impact of disease type has been reported for infliximab when used to treat Crohn's disease (58).

Other disease-related covariates would include ascites and pleural effusion. These comorbid conditions would be expected to increase the volume of distribution of proteins, thus lowering the measured concentrations in the serum. In one case (59), ascites was found to be a weak covariate of clearance as well.

41.2.2.3   Body Weight

As might be expected based on their mechanism of clearance, many monoclonal antibodies have demonstrated a strong relationship between weight and PK behavior (56). This relationship is often true for volume of distribution, since the distribution of most biologics is limited to extracellular fluid volume and one might expect that a patient with a high body weight would have a correspondingly larger extracellular fluid volume. Patient body weight can also be correlated to clearance in situations where receptor-mediated clearance is not predominant. The use of interspecies scaling has been shown to have some utility for scaling from animal to human exposure (60), although care should be taken to account for binding specificity and immunogenicity when attempting to scale the pharmacokinetics of human proteins evaluated in animal models.

41.2.2.4   Drug Interactions

The mixed function oxidases and cytochrome P450 enzyme systems do not play a role in the clearance of macromolecules. Nor do large proteins interact with transporter proteins such as P-gp, despite the fact that one site of clearance is the

intestines. Consequently, formal drug interaction studies are not often conducted for biologics. For example, there are several marketed therapeutic monoclonal antibodies (e.g., daclizumab, Zenapax®) that did not conduct formal drug interaction studies, although there are some marketed biologics that have reported drug interaction studies (61–63) with the expected negative outcomes. In general, reported drug interactions with biological agents and chemical agents are largely PD in nature (64, 65).

Although proteins are not cleared by cytochrome enzymes or mixed function oxidases, and would therefore not be expected to alter the pharmacokinetics of other medications, there are potential mechanisms of interaction between proteins and concomitant medications, which could affect patient exposure to the protein. For example, steroids, which alter macrophage cell trafficking (66, 67), could potentially alter the clearance of large therapeutic proteins.

An aspect of concomitant medications that should be considered during a population PK evaluation is previous treatment with other related biologic agents or others that have been derived from similar processes. For example, if a patient has developed antibodies to an agent that was derived from a prokaryotic cell line (e.g., *E. coli*), the patient may also be cross-reactive with a second protein that was derived from *E. coli* .

41.2.2.5   Liver Function

Proteins are not cleared by hepatic enzyme systems. However, liver size and function, spleen function, and macrophage function would be expected to account for variability in observed clearance of large therapeutic proteins. For instance, Sewell et al. (68) demonstrated that Kupffer cell function is decreased in aged rats, which is a function of age as much as it is of liver function. Standard measures of liver function such as alanine transferase may not provide relevant information and are rarely identified as a covariate even with small molecules. However, patients with advanced liver disease such as cirrhosis may have reduced clearance due to poor liver function and reduced liver blood flow.

41.2.2.6   Renal Function

Creatinine clearance is not usually a covariate for protein clearance. However, the kidney does form one site of clearance of proteins. Glomerular functionality and renal blood flow might be expected to have some impact on the clearance of low molecular weight proteins. For example, dose reductions are recommended in end stage renal disease patients receiving PEGylated interferon (52). However, accounts of reduced clearance in anuric patients are rare, suggesting that alternate routes of clearance are used in situations where renal function is nonexistent.

41.2.2.7   Age

As mentioned previously in Section 41.2.2.5, age can be a covariate for biologics. The effects of age on the PK properties of a protein appear to be due to changes in endothelial and macrophage function and to a lesser extent to changes in organ blood flow. Alterations of the immune response in elderly patients have been associated with increased amounts of memory and alloreactive T-cells, as well as altered cytokine responses (69), which can impact both on the pharmacokinetics and pharmacodynamics of a protein therapeutic agent.

41.2.2.8 Sex

After accounting for weight, patient sex is not usually identified as a covariate. However, as mentioned previously, there have been reports of differential expression of neutralizing antibodies between the two sexes (36). In addition, gender is occasionally identified. In one example, a population PK evaluation of cetuximab was performed (70). The covariates evaluated included demographic data (age, weight, height, body surface area, sex, and race), hepatic and renal function, cancer type, concurrent therapy, EGFr status, clinical response, and presence of a skin rash. A two-compartment model with saturable elimination was used to describe the concentration–time data. The volume of the central compartment was found to have a 27% reduction in the typical value of the central volume in females as compared to males and the typical value of V_{max} showed a 26% reduction in females, giving a maximal clearance from the saturable pathway of 0.059 L/h in males and 0.043 L/h in females. No other covariates were found to have a significant impact on the pharmacokinetics of cetuximab. In addition, enfuvirtide was found to have a 20% lower clearance in females than males, even after adjusting for body weight (52).

41.2.2.9 Race

Given the complexities of the pharmacology and pharmacokinetics of therapeutic proteins, the effect of race would not be expected to be important. There is no known difference between racial characteristics that would cause additional PK variability. Attempts have been made to examine the effect of race as a covariate, but it has only rarely been identified once patient weight or sex and other covariates (particularly those related to disease) were taken into account. Such was the case for cetuximab (70).

41.3 PHARMACODYNAMICS OF THERAPEUTIC PROTEINS: BACKGROUND

The PD response–time profiles for most of the biological agents on the market follow some variant of the basic indirect effect PD models described by Dayneka et al. (71). There are four basic indirect response models. The applicability of these models can be readily identified from the characteristic lag between plasma concentrations and measured response. In all four models described (71), there is a zero-order input rate constant of formation for the marker (K_{syn}) and a first-order degradation rate constant (K_{deg}). Prior to administration of a drug, the ratio of K_{syn} to K_{deg} determines the baseline level of the biomarker. An administered drug can act on either the rate of synthesis or rate of degradation and can be either inhibitory or stimulatory. The resulting change in biomarker levels depends on the site and type of action of the drug. For instance, an agent that acts to stimulate the synthesis rate constant would result in an increase in biomarker level. If that agent acted to inhibit the rate of synthesis, then the biomarker level would be expected to fall transiently. Once the effect of the drug wears off, the biomarker would be expected to return to the baseline level. The differential equation for an indirect system at the baseline state is given as

$$\frac{d\text{Biomarker}}{dt} = k_{\text{syn}} - k_{\text{deg}} \cdot A(1) \tag{41.6}$$

In this equation, Biomarker is the PD biomarker being described, k_{syn} is the synthesis rate constant of that biomarker, k_{deg} is the degradation rate constant for the biomarker, and $A(1)$ is the amount of biomarker. However, when drug is administered, the effect of that drug (*Eff* written below as a nonlinear stimulatory effect) can be added to the system (here it is added to the synthesis rate constant):

$$Eff = \frac{E_{\text{max}} \cdot C_p}{EC_{50} + C_p}$$

$$\frac{d\text{Biomarker}}{dt} = k_{\text{syn}} \cdot (1 + Eff) - k_{\text{deg}} \cdot A(1) \tag{41.7}$$

In this equation, *Eff* is the drug effect, E_{max} is the maximum effect of the drug, EC_{50} is the concentration of drug required to attain half-maximal effect, C_p is the concentration of the drug, Biomarker is the PD biomarker being described, k_{syn} is the synthesis rate constant of that biomarker, k_{deg} is the degradation rate constant for the biomarker, and $A(1)$ is the amount of biomarker.

If the delay between measured drug levels and response is very long, additional effect compartments can be added to allow the model to describe a longer lag period. A schematic diagram of an indirect effect model with the additional effect compartment for a precursor added is provided in Figure 41.3.

When the delay between concentration and response is protracted, additional effect compartments can be added to help describe the delay. These additional compartments also provide new places at which the drug effect can be evaluated.

The equations for this new schematic are given below. Note that K_{syn}, the rate of synthesis of the biomarker, has now become a first-order process and that the effect of drug has to be included for both the biomarker and the precursor pool.

$$Eff = \frac{E_{\text{max}} \cdot C_p}{EC_{50} + C_p}$$

$$\frac{d\,\text{Precursor}}{dT} = k_0 - k_{\text{syn}} \cdot A(\text{Precursor}) \cdot (1 + Eff) \tag{41.8}$$

$$\frac{d\text{Biomarker}}{dt} = k_{\text{syn}} \cdot A(\text{Precursor}) \cdot (1 + Eff) - k_{\text{deg}} \cdot A(\text{Biomarker})$$

FIGURE 41.3   Multiple compartment indirect effect model.

In this equation, *Eff* is the drug effect, E_{max} is the maximum effect of drug, EC_{50} is the concentration of drug required to attain half-maximal effect, C_p is the concentration of drug, Precursor is the predecessor for the biomarker, K_0 is the rate constant of formation of the precursor, Biomarker is the PD biomarker being described, K_{syn} is the synthesis rate constant of that biomarker, K_{deg} is the degradation rate constant for the biomarker, and $A(1)$ is the amount of biomarker.

This schematic provides new sites for adding a drug effect, such as on the rate of formation of the precursor pool (K_0). An additional aspect of this model is that it provides a basic mechanistic model for tolerance. If one assumes an agent works to stimulate K_{syn}, then the drug cannot have an effect when the amount in the precursor pool is depleted. Therefore, the duration of action of a drug can be dependent on the amount of precursor components that are available. There are many variants of this basic model, including the variation published by Movin-Osswald and Hammarlund-Udenaes (72) in 1995 and then by Sharma et al. (73) in 1998 in a slightly modified form.

Developing a PD model for a biological agent must be done on a case-by-case basis. The analyst must develop an understanding of the complex pharmacology that underlies the mechanisms of these agents. However, it should be noted that the basic PD behavior of many pharmacologically related proteins can often be described using similar models. For that reason, the PD behavior of several broad classes of therapeutic proteins may be broken down by the type of protein.

41.4   SPECIFIC PROTEINS

41.4.1   Cytokines

Cytokines form a family of proteins including interleukins and lymphokines that are released by cells in the immune system and act as intercellular mediators in immune response. Cytokines are produced by various cell populations, although they are predominantly produced by helper T cells and macrophages. Cytokines that are secreted from lymphocytes are referred to as lymphokines, whereas those secreted by monocytes or macrophages are referred to as monokines. Many lymphokines are also referred to as interleukins (ILs), because they are also capable of affecting leukocyte cellular responses.

Several different broad classes of cytokines are produced by the body, the largest of which stimulates immune cell proliferation and differentiation. The largest class includes IL-1, which activates T cells; IL-2, which stimulates proliferation of antigen-activated T and B cells; IL-4, IL-5, and IL-6, which all stimulate proliferation and differentiation of B cells; interferon gamma (IFN-γ), which activates macrophages; and IL-3, IL-7, granulocyte-macrophage colony-stimulating factor (GM-CSF), and granulocyte colony-stimulating factor (G-CSF), all of which stimulate hematopoiesis.

There are not many examples of cytokines that have been approved as therapeutic agents. In part, this lack is due to the pleiotropic effects of these agents (74), which makes the overall effect difficult to predict, and because administration of these agents will often mediate increased immunological response. In general, the administration of interleukins results in elevated blood cell counts, particularly white cells. However, because they help to increase blood cells and also induce

immune response, interleukins have been found to be useful in the treatment of advanced cancers. Interleukin-2 (Proleukin®, aldesleukin, IL-2) was approved for treatment of metastatic renal cell carcinoma in 1992, and then later for the treatment of metastatic melanoma in 1998.

The pharmacokinetics and pharmacodynamics of recombinant interleukin-2 (IL-2) in patients with human immunodeficiency virus (HIV) infection have been evaluated (75). Patients were administered IL-2 either by continuous infusion or by SC injection for 5 days over multiple cycles. Following repeated injection, soluble IL-2 receptors were substantially but transiently increased. A dose-dependent decrease in area under the concentration–time curve (AUC) between days 1 and 5 was attributed to a receptor-mediated change in clearance. Concentrations were described using an unusual model that employed an indirect stimulatory PD model to link the time-dependent changes of the pharmacokinetics with the change in IL-2 receptor density following repeated administration.

41.4.2 Interferons

Interferons, which were discovered in the 1950s as a result of their antiviral activity (76), are pleiotropic agents exhibiting a wide variety of effects including antiviral, antiproliferative, hematopoietic, and immunomodulatory activities (77, 78). Interferons are sometimes considered to be cytokines because of their role in cellular and humoral immune responses.

Interferons are generally stimulatory proteins that exert their activity through interactions with cell surface receptors, inducing cellular processes and enhancing specific gene translation (79). Interferons also regulate the expression of unique antiviral proteins such as MX protein, which alters microtubule formation and mitosis, and $2'$-$5'$-oligoadenylate synthetase (2,5-OAS), which induces the destruction of viral RNA.

A general schematic diagram of the mechanism of action of interferons on cellular protein production is presented in Figure 41.4. The pharmacodynamics of interferon alpha using MX protein as the biomarker have been described using a simple indirect effect stimulatory model (55).

Although they predominantly exhibit stimulatory activity, interferons can inhibit general cellular protein synthesis, including the synthesis of cytochrome P450 enzymes, making interferon one of the few biological agents that have the potential for causing "classic" drug interactions.

The mechanism by which IFN-α exerts antitumor activity is unclear, particularly in hematological cancers. In melanoma and renal cell carcinoma, antitumor effects may be mediated by augmented immune responses including activation of natural killer lymphocytes and enhanced expression of cell surface antigens (e.g., MHC I and II). However, these mechanisms have not been decisively proved.

41.4.3 Growth Factors

Growth factors are proteins that bind to receptors on the cell surface, with the primary result of activating cellular proliferation and/or differentiation. Many growth factors are quite versatile, stimulating cellular division, maturation, and margination in numerous different cell types; while other factors are specific to

FIGURE 41.4   Mechanism of action of interferon on cellular processes. Interferon (IFN) binds to a cell surface receptor, which, through a series of cellular processes, enhances the formation of RNA and subsequently increases protein formation when the RNA binds to ribosomes (hexagons). The proteins (cylinder) can either remain in the cell or be excreted.

a particular cell type. Growth factors are not the only agents that exhibit hematopoietic activity; cytokines are also capable of modulating cell growth and maturation.

At present, there are several hematopoietic factors that have been approved for clinical use. These approved agents include granulocyte growth and stimulation factor (G-CSF, Filgrastim, Neupogen®) and its PEG-modified variant (Neulasta®), granulocyte-macrophage colony-stimulating factor (GM-CSF), stem cell factor (SCF), and erythropoietin (EPO) and its hyper-glycosylated variant (Aranesp®). A simplified diagram of the hematopoiesis is provided in Figure 41.5. The growth factors and cytokines that influence each pathway are provided in the figure.

One aspect that must always be considered in the development of any PD model for a biological agent is the presence of endogenous factors that will also influence the measured biomarker activity. This is particularly true for growth factors. For example, patients with neutropenia following chemotherapy will have elevated endogenous levels of G-CSF. The same is true of an anemic patient, whose endogenous EPO levels will be elevated. The pharmacodynamics of growth factors are complex, but they do have some common characteristics between classes.

41.4.4   G-CSF

G-CSF is a protein that regulates the production of neutrophils by stimulating neutrophil progenitor proliferation (80, 81), differentiation (82), and selected end-cell functional activation (83). G-CSF has little effect on the production of other hematopoietic cells. Endogenously, G-CSF is produced by monocytes, fibroblasts, and endothelial cells. As a low molecular weight protein (MW < 20 kD), G-CSF is subject to clearance by glomerular filtration and there is good evidence that

FIGURE 41.5 Schematic diagram of hematopoiesis. The cytokines and other factors that control cell formation and maturation are provided for reference.

endogenous concentrations are also cleared by binding to a receptor on the surface pluripotent stem cells as well as neutrophils (84). The first binding effects the differentiation of the stem cells toward eventual maturation to neutrophils. The latter receptor interaction appears to play a critical role in hemostasis, increasing clearance of G-CSF from both endogenous and exogenous sources when cell counts are high as a means of controlling neutrophil count by a feedback mechanism (31). Because G-CSF works to increase circulating neutrophils, the clearance of G-CSF varies over the course of treatment and is dependent on individual PD response (85). There is a strong PK–PD interaction with this agent and, therefore, a physiological limitation to the PD activity of G-CSF. Wahlby et al. (86) have demonstrated the importance of using time-dependent covariates and this is particularly relevant with biological agents.

Another aspect of the pharmacological response to G-CSF administration is the unusual pattern of induced changes in neutrophil count over time. Immediately following the first administration, G-CSF dose-independently induces neutropenia

and causes substantial downregulation of its own receptor (CD114) on neutrophils (87). This G-CSF–CD114 interaction dose-independently induces degranulation of neutrophils, which results in increased levels of gelatinase B, an enzyme that precipitates neutropenia and subsequent neutrophilia. The gelatinase B release into plasma may also contribute to mobilization of neutrophils or stem cells into peripheral circulation.

The PK–PD relationship for G-CSF following IV and SC administration was well characterized in healthy volunteers (53). The PK model was a two-compartment PK model with bisegmental absorption from the site of SC administration, parallel first-order and saturable elimination pathways, and an indirect effect PD model describing the time course of neutrophils. A sigmoidal E_{max} model was applied for the stimulation of the neutrophil input rate. In addition, a time-variant scaling factor for absolute neutrophil count (ANC) observations was introduced to account for the early transient depression of ANC.

A simple indirect effect stimulatory model adequately describes the time course of neutrophils following G-CSF administration. However, G-CSF is commonly administered following chemotherapy to treat the associated neutropenia. There is a substantial lag time between the administration of chemotherapeutic agents and the nadir ANC value that can be described more accurately using the "cell transit" PD model (88) than a simple indirect effect model. The cell transit PD model utilizes a gamma distribution to provide a semiphysiological description of cell maturation. The schematic diagram for this model is provided in Figure 41.6.

Despite its apparent complexity, this model is relatively easy to use. There is only one transit rate constant "Kt" that is used to describe the transfer of cells from one compartment to the next. Ksyn and Kdeg are the synthesis and degradation rate constants, respectively. Chemotherapy is assumed to act on the synthesis rate constant in an inhibitory fashion. The effect of G-CSF can also be added to this model, making it a better model for comparing the efficacy of G-CSF and other variants in a clinically relevant system.

41.4.5   EPO

Unlike G-CSF, erythropoietin (EPO) is a pleiotropic agent with multiple actions and different sites of activity. The various sites that EPO can affect in the hema-

FIGURE 41.6   Schematic diagram of "cell transit" model. In this model, pluripotent stem cells are assumed to be produced on a first-order process, and differentiation (i.e., passing though each stage or compartment of cell growth) occurs over a fixed transit time (Kt). Mature cells are lost via a first-order process (Kdeg). The model can also allow for a negative feedback by which mature white cells diminish the formation of new stem cells.

topoietic chain are shown in Figure 41.5. Primarily, EPO functions to maintain appropriate oxygenation of cells and its production is regulated via feedback from oxygen pressure detecting cells in the kidneys. A schematic diagram of the process of red blood cell (RBC) production and maturation is provided in Figure 41.7. In the circulatory system, EPO also effects the development of new blood vessels.

EPO may also act to facilitate the survival and proliferation of nonerythroid cells as well. In addition to production in the kidneys, EPO is produced in the brain, although this form has a lower molecular weight than the peripherally produced variant (89). In the central nervous system (CNS), EPO plays a critical role in brain function and development. EPO receptors have also been isolated in ovary, oviduct, uterine, and testes cells (90), as well as some tumor cell lines such as breast cancer (91). The function of EPO binding in these alternate cell types has not yet been determined, although the appearance of EPO receptors on cancer cells has been indicated as a poor prognosis factor (92). The use of EPO to treat anemia arising from chemotherapy or from cancer has therefore been called into question (93).

Recombinant human EPO has a relatively low molecular weight (MW = 30.4 kD). Because of its low molecular weight, EPO would be expected to undergo both renal and hepatic elimination, although these routes of elimination appear to be relatively minor (94). The pharmacokinetics of EPO have been extensively studied, and in many cases, nonlinear elimination was reported (95) and the clearance was determined to change following phlebotomy or bone marrow ablation in a fashion that is consistent with receptor-mediated clearance. Following SC administration to healthy volunteers, the pharmacokinetics of EPO were best described with a dual-absorption rate model (fast zero-order and slow first-order inputs) with nonlinear disposition (96).

Unlike white cells, which are generally assumed to follow first-order kinetics, red cells are attributed as having a lifespan of 120 days in a normal adult (97).

FIGURE 41.7   Mechanism for control of endogenous erythropoietin and red cell production. The regulation of EPO production depends on the oxygenation of the blood as it passes through the kidney. If the oxygen levels are low, the kidney synthesizes additional EPO, which acts to prolong the survival of the blast forming units (BFU-E) and colony forming units (CFU-E), allowing enhanced red blood cell (RBC) production.

This theory was originally established in the 1960s and still remains in place today. Therefore, a specialized set of indirect PD models for agents that alter the generation of natural cells based on a lifespan concept were developed (98) based on the concept of a fixed lifespan.

In this "lifespan model," mature cells are assumed to be produced at a constant rate and to survive for a fixed interval of time, after which they are lost. Therefore, rate of cell loss must equal the cell production rate at the time those cells were produced (e.g., one "lifespan" ago). This aspect of lifespan-mediated loss requires that the model also track the number of cells produced at a prior interval of time. A stimulatory or inhibitory effect of a therapeutic agent such as EPO then results in a delayed increase in cell count (depending on the number of intervening compartments) but the return to baseline cell count is dependent on the lifespan of the cells.

It is interesting to note that the "transit model" and the "lifespan model" both produce a very similar time course of effect if the transit model has sufficient (e.g., at least five) compartments.

41.4.6 Antibodies

When engineered monoclonal antibodies were initially undergoing development as therapeutic agents, it was assumed that the mechanism of activity of these antibodies was directly related to the effector activity of antibodies in vivo. That is, the MAb would bind to its target receptor, which would precipitate antibody-dependent cell cytotoxicity (ADCC) by attracting natural killer (NK) cells. Because many of these early MAbs were targeted against cancer cells, ADCC was considered a desirable mechanism of action, although these early theories have not been borne out.

NK cells provide two types of effector function: cell cytotoxicity and lymphokine secretion. In conjunction with antibodies, NK cells can cause cytotoxicity through recognition and lysis of MAb-coated target cells. A proposed schematic for the mechanism of such "lytic" antibodies is presented in Figure 41.8. NK cells also possess a wide range of regulatory receptors that can prevent cytotoxic responses (99) based on the cell surface expression of killer-cell inhibitory receptors (KIRs). Regulation of NK cell effector activity has called into question the extent to which

FIGURE 41.8   Mechanism of action of "lytic" monoclonal antibodies. MAb binds to receptor on antigen presenting cell. The receptor binding attracts phagocytes, resulting in cell destruction.

therapeutic MAbs utilize cytotoxic effector mechanisms to provide clinical benefit. It is also unclear whether the concentrations of MAbs following therapeutic dosing are sufficiently high to saturate the substantial numbers of targeted receptors present in many therapeutic indications. Conversely, there is evidence from Fc receptor knockout mice suggesting that, in certain systems, Fc binding is required for clinical activity of therapeutic MAbs (100). Furthermore, anticancer MAbs do not appear to have functional activity when administered as a Fab fragment, which would support the theory that clinical activity of MAbs is linked to effector function. Fc receptors do more than recruit effectors; therefore, part of the requirement for Fc function may be attributable to crosslinking on the target cells, which interferes with cellular function.

However, many therapeutic MAbs are currently being investigated for autoimmunity and immunosuppression in therapeutic areas such as rheumatoid arthritis, where the role of ADCC is less appropriate (101). An alternative mechanism of blocking or modulating responses is more desirable than ADCC. A key aspect of this mechanism is that a relatively short exposure to MAb may break the inflammatory cycle and allow the repair process to begin. A schematic diagram of this mechanism is presented in Figure 41.9. MAbs therefore can provide a long-term effect following short-term treatment through a mechanism referred to as "infectious tolerance" (102).

The PD behavior of therapeutic MAbs is complex. Fortunately, the activity of these agents can be evaluated by the use of fluorescence activated cell sorting (FACS). Like most therapeutic biologics, the mechanism of action is commonly described using the standard indirect effect models. However, because FACS data can determine the fraction of cell surface receptor bound by the MAb, and because the change in receptor density can be followed using this same method, more mechanistic models have been proposed (29). In these models, the relationship between concentration of antibody and bound receptors can be explicitly described and the bound antibody is then used to drive the indirect effect model. A schematic of this model is provided in Figure 41.10.

FIGURE 41.9 Mechanism of action of coating monoclonal antibodies. MAb binds to receptor on antigen presenting cell. The receptor binding is "nonproductive", resulting in a stimulation of receptor loss. Alternatively, when the MAb binds to the receptor, it results in a steric blockage of receptor.

FIGURE 41.10   Example schematic for PK/PD model for therapeutic monoclonal antibodies. The antibody binds (potentially reversibly) to the receptor on the cell surface. The bound antibody–receptor complexes are then what drives the stimulation of loss of receptor-positive (R+) cells.

It should be noted that, unlike small molecules, the action of MAbs is often directly related to the bound concentration of drug rather than the free concentrations, in contrast to the PD behavior of small molecules.

41.5   COVARIATES FOR PHARMACODYNAMIC RESPONSE

When modeling the pharmacodynamics of a biological agent, some consideration needs to be given to the identification of covariates. Because the physiological system that is targeted by the agent is often well characterized, the selection of covariates for investigation can often be limited to those that have a strong likelihood of being identified. For instance, when characterizing the pharmacodynamics of a hematopoietic factor, the time to last treatment with chemotherapy or even the type or number of cycles of treatment might be projected to be covariates. Chemotherapy results in loss of bone marrow function, which in turn should reduce the PD response in a patient. Similarly, concomitant administration of drugs that alter receptor density could also affect the PD response of the investigational agent. PD models often take a long time to converge, so the number of covariates that can be investigated is often limited. Careful selection of the covariates for evaluation is therefore necessary.

41.6   EVALUATING PHARMACODYNAMICS USING MODEL-BASED ANALYSIS

Three examples of a NONMEM (Version 5, Globomax LLC, Hanover, MD) control stream and the necessary data format for a commonly employed PK/PD model are

provided in the following section. The generic code and associated example data file format must be suitably altered for more complex models.

41.6.1 Control Stream

One of the more common PD models used to describe the time course of a biomarker is the simple indirect effect model. In this example, drug concentrations increase the rate of degradation of the biomarker and act to reduce the biomarker concentration. Such a model has been used to characterize the pharmacokinetics/pharmacodynamics of PEGylated interferon (55) and other proteins. A fragment of an example control stream for this basic PK/PD model is presented and explained in Table 41.5. Explanations for the different commands used are provided as well.

TABLE 41.5   Example NONMEM Code 1: Commonly Used Indirect Effect PK/PD Model

NONMEM Code	Explanation
```$SUBROUTINES ADVAN6``` ``` TRANS1 TOL 3```	Typically evaluating a PK/PD model requires the use of ADVAN6 (or one of the other ADVANs used to evaluate differential equations). Initially, TOL is set to a low value such as 3 to facilitate convergence.

```$MODEL```
```COMP=(CENTRAL,DEFOBS)```
```COMP=(PERIPH)```
```COMP=(EFFECT)```
	These lines of code define the PK and PD model compartments. Here, the first two compartments are defined for the pharmacokinetics and the third compartment is for the biomarker.

```$PK```
```;Define the PK Parameters```
```CALLFL=-22```
```TVMX = THETA(1)```
```VMAX = TVMX*EXP(ETA(1))```
```TVKM = THETA(2)```
```KM=TVKM*EXP(ETA(2))```
```K23 = THETA(4)*EXP(ETA(4))```
```K32 = THETA(5)*EXP(ETA(5))```
```TVV1 = THETA(3)```
```V1 = TVV1*EXP(ETA(3))```
	This portion of the code defines the parameters for the PK model. The semicolon is a "comment" statement and is not read by NONMEM. Whenever developing a model, adding good commenting is always recommended. This is particularly true when the model is complex and the control stream is long.
	The present PK model is a two-compartment model with Michaelis–Menton elimination. VMAX and KM define the parameters for elimination, K23 and K32 define the intercompartmental transfer rate constants, and V1 is the central volume of distribution. As with many biologics, this theoretical agent is being administered intravenously.

```;DEFINE PD PARAMS```
```TVKDEG=THETA(6)```
```KDEG=TVKDEG```
```TVKSYN=THETA(7)```
```KSYN=TVKSYN*EXP(ETA(8))```
```TVEMAX=THETA(8)```
```EMAX=TVEMAX*EXP(ETA(6))```
```TVEC50=THETA(9)```
```EC50=TVEC50*EXP(ETA(7))```
	This portion of code defines the parameters for the PD model. This is a simple indirect effect stimulatory model. KDEG is the degradation rate constant for the biomarker, KSYN is the formation rate constant of the biomarker, EMAX is the maximal effect of the drug, and EC50 is the concentration at which half-maximal effect occurs.

TABLE 41.5 *Continued*

NONMEM Code	Explanation
| ```
;SCALE COMPARTMENTS
S1=V1
F3=KSYN/KDEG
``` | This section of code scales the PK compartments. An important aspect to note here is the use of F3, which is the bioavailability term for the effect compartment. Here, F3 is used to initialize the effect compartment to the baseline value of the biomarker being modeled, which is why it is set to the ratio of KSYN/KDEG. Prior to therapy, the biomarker is presumed to be at some steady-state value, which should be equivalent to the ratio of formation to degradation. F3 is used in conjunction with a special dose item for this compartment to initialize the effect compartment. |
| ```
$DES
;PK Model
  CP=A(1)/V1
  ;PLASMA CONCS IN UG/L
  CLMM=(CP*VMAX)/(KM+CP)
  DADT(1)=-CLMM-K23
    *A(1)+K32*A(2)
  DADT(2)=K23*A(1)-K32*A(2)
;PD Model
  EFF=EMAX*CP/(EC50+CP)
  DADT(3)=KIN-KOUT*A(3)
    *(1+EFF)
``` | This section contains the differential equations that define the pharmacokinetics and the pharmacodynamics of the drug. Here, the effect is stimulatory on KDEG, which will result in a transient reduction of the biomarker. |
| ```
$ERROR
QK=0
OD=0
IF (CMT .EQ. 1) QK=1
IF (CMT .EQ. 3) QD=1
  PKY = F * EXP(ERR(1))
    + ERR(2)
  PDY = F+ERR(3)
Y=QK*PKY+QD*PDY
``` | This section defines the residual error models for the pharmacokinetics and pharmacodynamics. Note that the PD observations have a simple additive error function, which is different from the PK residual error function. Residual error models must be selected separately for the PK and PD models. |

It is important to realize that the parameter estimates obtained for one agent do not usually translate readily to other agents for the PD models. Furthermore, the evaluation of complex sets of differential equations using NONMEM generally requires that the initial estimates for parameters be reasonable since the models can find local minima or may not converge at all if the initial estimates for the parameters are not reasonable. Therefore, when developing a model for a new agent, care should be taken to ensure that the initial estimates are good, and running the model without the estimation step to evaluate the performance of these estimates is critical.

A more complex binding type model is defined in Table 41.6. The model defined here is similar to the model shown in Figure 41.10, although it uses scaled binding

TABLE 41.6   Example NONMEM Code 2: "Cell Transit" Gamma Distribution Model

```
$SUBS ADVAN6 TOL=3
$MODEL
  COMP=(CENTRAL)
  COMP=(PERIPH)
  COMP=(STEM)
  COMP=(WBC)
  COMP=(TRANSIT1)
  COMP=(TRANSIT2)
  COMP=(TRANSIT3)
"FIRST
"COMMON/PRCOMG/IDUM1,IDUM2,IMAX,IDUM4,IDUM5
"INTEGER IDUM1,IDUM2,IMAX,IDUM4,IDUM5
"IMAX=50000000
$PK
  V1  = THETA(4)*EXP(ETA(2))
  K10 = THETA(1)*EXP(ETA(1))
  K12 = THETA(2)
  K21 = THETA(3)
  BASE = THETA(5)*EXP(ETA(3))
  IF (BASE .LE. 0) EXIT 1 101
  MTT = THETA(6)*EXP(ETA(4))
  K = 4/MTT
  F3 = BASE
  F4 = BASE
  F5 = BASE
  F6 = BASE
  F7 = BASE
  POWER = THETA(7)
  SLOP = THETA(8)*EXP(ETA(5))
$DES
;Pharmacokinetics
CP = A(1)/VOF
DRUG = SLOP*CP
;***************KINETICS***************************
DADT(1)=A(2)*K21-A(1)*K10-A(1)*K12
DADT(2)=A(1)*K12-A(2)*K21
;***************DYNAMICS***************************
DADT(3) = K*A(3)*(1-DRUG)*(BASE/A(4))**POWER - K*A(3)  ;Pharmacodynamics
DADT(4) = K*A(7) - K*A(4)    ;WBC
DADT(5) = K*A(3) - K*A(5)    ;TRANSIT 1
DADT(6) = K*A(5) - K*A(6)    ;TRANSIT 2
DADT(7) = K*A(6) - K*A(7)    ;TRANSIT 3
$ERROR
IF (CMT .EQ. 1) QK=1
IF (CMT .EQ. 4) QD=1
PKY = F * EXP(ERR(1)) + ERR(2)
PDY = F * EXP(ERR(3)) + ERR(4)
Y=QK*PKY+QD*PDY
```

kinetics to define the binding. This control stream has several similarities to the previous example in that a stimulatory indirect effect model is used to describe the change in time course of the biomarker, although in this present example, it is bound drug causing the change in biomarker, not the total drug concentration as was seen previously. This control stream uses a separate compartment for the binding of drug, with limitations set to scale the amount of drug that can bind. Information to develop these models must usually combine information from in vitro studies as well as nonclinical pharmacology studies.

Table 41.6 shows the control stream used to implement the model shown in Figure 41.6. This is the "cell transit" model, which is defined by a simple queuing or gamma function. The model is particularly useful for describing the pharmacodynamics of agents that impact cell growth or turnover or have very long lag times.

41.6.2   Database Requirements

An example database for a two-compartment PK model driving a simple indirect effect model is presented in Table 41.7. NONMEM nominally assigns a value of 0

TABLE 41.7   Example Database Format for Indirect Effect Model

CID	DATE	TIME	AMT	RATE	DV	CMT	MDV	EVID
1	7/16/2000	8:00	1	.	.	3	1	1
1	7/31/2000	13:20	3,000	1,500	.	1	1	1
1	7/31/2000	13:20	.	.	124	3	.	.
1	8/2/2000	10:15	30,000	15,000	.	1	1	1
1	8/2/2000	12:47	.	.	700	1	.	.
1	8/7/2000	10:45	30,000	15,000	.	1	1	1
1	8/7/2000	10:45	.	.	37	3	.	.
1	8/7/2000	13:15	.	.	729	1	.	.
2	12/19/1991	2:00	1	.	.	3	1	1
2	1/3/1992	8:00	.	.	6	3	.	.
2	2/3/1992	15:00	.	.	8	3	.	.
2	2/3/1992	15:15	7,500	3,600	.	1	1	1
2	2/6/1992	16:59	.	.	7	3	.	.
2	2/10/1992	8:50	7,500	3,600	.	1	1	1
2	2/10/1992	10:05	.	.	5	3	.	.
2	2/13/1992	0:00	.	.	6	3	.	.
2	2/17/1992	8:36	.	.	6	3	.	.
3	12/25/1991	8:00	1	.	.	3	1	1
3	1/9/1992	11:06	.	.	117	3	.	.
3	1/27/1992	8:21	.	.	176	3	.	.
3	1/27/1992	11:30	7,500	3,600	.	1	1	1
3	1/27/1992	11:43	.	.	155	3	.	.
3	1/27/1992	14:30	.	.	81	3	.	.
3	1/27/1992	16:30	.	.	91	3	.	.
3	1/27/1992	18:30	.	.	114	3	.	.
3	2/2/1992	9:15	.	.	169	3	.	.

for the initial conditions for each compartment. While the assumption that the initial condition is 0 is acceptable for a PK model, it is not an appropriate initial condition for many PD models. This is particularly true for indirect effect models because the biomarker is usually present at some steady-state level prior to the administration of drug. Once drug is administered, the biomarker will either increase or decrease depending on the activity of the drug. Therefore, it is necessary to use a bioavailability term to initialize the effect compartment and to provide a unit dose into that compartment in the data set. The unit dose record should be read in prior to the other records so that the effect compartment can be initialized. It need be entered only once per individual.

The other specialty item for an indirect PK/PD database is the use of the CMT item. This item is set to 1 for the PK observations and to 3 for the PD observations in the example data provided so that the observations are properly associated with their assigned compartments. In this example database, the PK model was assumed to be a two-compartment model. For more information on the use of special data items, see the *NONMEM User Guides* (103).

41.7   SUMMARY

Developing models that describe the pharmacokinetics and pharmacodynamics of therapeutic proteins is a challenging process. Before getting started, it is always best to develop an understanding of the system that the therapeutic protein will impact. It is also helpful to review available information on the PK behavior of other related proteins. Understanding the design and manufacture of the biological agent is also critical. For instance, in many cases, proteins that have been modified by the addition of such agents as polyethylene glycol will have lower binding affinities to target receptors, but these agents also have a much lower systemic clearance, yielding a net therapeutic benefit.

During evaluation of these agents, the usual covariates that are evaluated for chemically based drugs such as creatinine clearance may not be relevant given the size and modification of the biologic. Careful consideration of the protein and its pharmacology is helpful to determine what covariates are likely to be relevant. Similarly, it is not uncommon for the PD activity of a therapeutic protein to have an impact on the pharmacokinetics. This is particularly true for agents that are cleared by binding to a receptor and, through that binding, alter the receptor expression.

In most cases, the PD activity of a therapeutic protein does not follow direct effect type behavior. Rather, these agents act at a cellular level and the resulting activity is governed by slower moving processes such as cell or protein turnover. Again, understanding the pharmacology of the therapeutic agent will be helpful in designing a PD model that adequately describes the activity of the drug.

The information provided in this chapter is meant to serve only as a basic overview to therapeutic proteins. There are many agents that have been developed that cannot be described using the basic concepts described here. Evaluation of these agents is always a learning process!

ABBREVIATIONS

ADCC	Antibody-dependent cell cytotoxicity
ANC	Absolute neutrophil count
CFU	Colony forming units
CHO	Chinese hamster ovary
EPO	Erythropoietin
Fab	Fragment antigen binding (binding region of a monoclonal antibody)
FACS	Fluorescence activated cell sorting
FC	Fragment constant (heavy chain region of monoclonal antibody)
G-CSF	Granulocyte colony-stimulating factor
HAGA	Human anti-globulin antibody
HAMA	Human anti-murine antibodies
IFN	Interferon
IG	Immunoglobin
IL	Interleukin
K	Rate constant
kD	Kilodaltons (a molecular weight for proteins)
KIR	Killer-cell inhibitory receptor
MAb	Monoclonal antibody
MW	Molecular weight
NAb	Neutralizing antibody
NK	Natural killer
N-NAb	Nonneutralizing antibody
PD	Pharmacodynamic
PEG	Polyethylene glycol
PK	Pharmacokinetic
PMNs	Polymorphonuclear neutrophils
R+	Receptor positive
RBC	Red blood cell
rHu	Recombinant human
SC	Subcutaneous
WBC	White blood cell

REFERENCES

1. M. L. Rodrigues, B. Snedecor, C. Chen, W. L. Wong, S. Garg, G. S. Blank, D. Maneval, and P. Carter, Engineering Fab fragments for efficient F(ab)2 formation in *Escherichia coli* and for improved in vivo stability. *J Immunol* **151**(12):6954–6961 (1993).

2. P. Tanswell, N. Modi, D. Combs, and T. Danays, Pharmacokinetics and pharmacodynamics of tenecteplase in fibrinolytic therapy of acute myocardial infarction. *Clin Pharmacokinet* **41**(15):1229–1245 (2002).

3. G. Kohler and C. Milstein, Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**(5517):495–497 (1975).

4. J. J. Tjandra, L. Ramadi, and I. F. McKenzie, Development of human anti-murine antibody (HAMA) response in patients. *Immunol Cell Biol* **68**(Pt 6):367–376 (1990).

5. G. L. DeNardo, B. M. Bradt, G. R. Mirick, and S. J. DeNardo, Human antiglobulin response to foreign antibodies: therapeutic benefit? *Cancer Immunol Immunother* **52**(5):309–316 (2003).

6. G. L. Boulianne, N. Hozumi, and M. J. Schulman, Production of functional chimaeric mouse/human antibody region domains. *Nature* **312**:643–646 (1984).

7. S. L. Morrison, M. J. Johnson, L. A. Herzenberg, et al., Chimeric human antibody molecules: mouse antigen-binding domains with human constant. *Proc Natl Acad Sci USA* **81**:6851–6855 (1984).

8. J. D. Isaacs, The antiglobulin response to therapeutic antibodies. *Semin Immunol* **2**:449–456 (1990).

9. A. Abuchowski, J. R. McCoy, N. C. Palczuk, T. van Es, and F. F. Davis, Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *J Biol Chem* **252**(11):3582–3586 (1977).

10. J. M. Harris and R. B. Chess, Effect of PEGylation on pharmaceuticals. *Nat Rev Drug Discov* **2**(3):214–221 (2003).

11. J. M. Harris, N. E. Martin, and M. Modi, PEGylation: a novel process for modifying pharmacokinetics. *Clin Pharmacokinet* **40**(7):539–551 (2001).

12. A. Cases, Darbepoetin alfa: a novel erythropoiesis-stimulating protein. *Drugs Today (Barc)* **39**(7):477–495 (2003).

13. M. Allon, K. Kleinman, M. Walczyk, C. Kaupke, L. Messer-Mann, K. Olson, A. C. Heatherington, and B. J. Maroni, Pharmacokinetics and pharmacodynamics of darbepoetin alfa and epoetin in patients undergoing dialysis. *Clin Pharmacol Ther* **72**(5):546–555 (2002).

14. S. Elliott, T. Lorenzini, S. Asher, K. Aoki, D. Brankow, L. Buck, L. Busse, D. Chang, J. Fuller, J. Grant, N. Hernday, M. Hokum, S. Hu, A. Knudten, N. Levin, R. Komorowski, F. Martin, R. Navarro, T. Osslund, G. Rogers, N. Rogers, G. Trail, and J. Egrie, Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat Biotechnol* **21**(4):414–421 (2003).

15. A. I. Fernandes and G. Gregoriadis, The effect of polysialylation on the immunogenicity and antigenicity of asparaginase: implication in its pharmacokinetics. *Int J Pharm* **217**:215–224 (2001).

16. T. A. Waldman et al., Variations in the metabolism of immunoglobulins measured by turnover rates, in *Immunoglobulins: Biological Aspects and Clinical Uses*, E. Merler (Ed.). National Academy of Sciences, Washington, DC, 1970, p.38.

17. D. Colcher, R. Bird, M. Roselli, K. D. Hardman, S. Johnson, S. Pope, S. W. Dodd, M. W. Pantoliano, D. E. Milenic, and J. Schlom, In vivo tumor targeting of a recombinant single-chain antigen-binding protein. *J Natl Cancer Inst* **82**:1191–1197 (1990).

18. J. A. Marinaro, D. J. Casley, and L. A. Bach, O-glycosylation delays the clearance of human IGF-binding protein-6 from the circulation. *Eur J Endocrinol* **142**(5):512–516 (2000).

19. T. W. Hepburn, T. K. Hart, V. L. Horton, T. S. Sellers, L. P. Tobia, J. J. Urbanski, W. Shi, and C. B. Davis, Pharmacokinetics and tissue distribution of SB-251353, a novel human CXC chemokine, after intravenous administration to mice. *J Pharmacol Exp Ther* **298**(3):886–893 (2001).

20. T. Maack, Renal handling of low molecular weight proteins. *Am J Med* **58**(1):57–64 (1975).

21. F. A. Carone, D. R. Peterson, S. Oparil, and T. N. Pullman, Renal tubular transport and catabolism of proteins and peptides. *Kidney Int* **16**(3):271–278 (1979).

22. B. M. Brenner, *Brenner's The Kidney,* 7th ed. Saunders Publishing, Philadelphia, PA, 2004, Vol. 1, pp. 353–452.

23. W. R. Clark and D. Gao, Low-molecular weight proteins in end-stage renal disease: potential toxicity and dialytic removal mechanisms. *J Am Soc Nephrol* **13**:S41–S47 (2002).

24. S. Yokota and H. D. Fahimi, Uptake of formalin-denatured albumin by the sinus-lining cells of rat liver: an immunocytochemical and cytochemical study. *Cell Struct Funct* **12**(3):295–309 (1987).

25. A. Joshi, Monoclonal antibody roundtable. AAPS Annual Meeting, Denver, CO, November 2003.

26. A. Avhandling, *Elimination of Circulating IgA and IgG Immune Complexes, An Experimental Study in Mice, Rats and Guinea Pigs*. Linkoping University Medical Dissertation No. 474, 1995.

27. A. K. Abbas and A. H. Lichtman, *Cellular and Molecular Immunology*, 4th ed. Saunders Publishing, Philadelphia, PA, 2000, pp. 25–26.

28. E. K. Rowinsky, G. H. Schwartz, J. A. Gollob, J. A. Thompson, N. J. Vogelzang, R. Figlin, R. Bukowski, N. Haas, P. Lockbaum, Y. P. Li, R. Arends, K. A. Foon, G. Schwab, and J. Dutcher, Safety, pharmacokinetics, and activity of ABX-EGF, a fully human anti-epidermal growth factor receptor monoclonal antibody in patients with metastatic renal cell cancer. *J Clin Oncol* **22**(15):3003–3015 (2004).

29. D. R. Mould, C. B. Davis, E. A. Minthorn, D. C. Kwok, M. J. Eliott, M. E. Luggen, and M. C. Totoritis, Population pharmacokinetic/pharmacodynamic analysis of the effects of Clenoliximab, a PRIMATIZED™ anti-CD4 monoclonal antibody, on T lymphocytes, following single doses to patients with active rheumatoid arthritis. *Clin Pharmacol Ther* **66**(3):246–257 (1999).

30. S. G. Ericson, H. Gao, G. H. Gericke, and L. D. Lewis, The role of polymorphonuclear neutrophils (PMNs) in clearance of granulocyte colony-stimulating factor (G-CSF) in vivo and in vitro. *Exp Hematol* **25**(13):1313–1325 (1997).

31. K. Terashi, M. Oka, S. Ohdo, T. Furukubo, C. Ikeda, M. Fukuda, H. Soda, S. Higuchi, and S. Kohno, Close association between clearance of recombinant human granulocyte colony-stimulating factor (G-CSF) and G-CSF receptor on neutrophils in cancer patients. *Antimicrob Agents Chemother* **43**:21–24 (1999).

32. D. E. Mager, M. A. Mascelli, N. S. Kleiman, D. J. Fitzgerald, and D. R. Abernethy, Simultaneous modeling of abciximab plasma concentrations and ex vivo pharmacodynamics in patients undergoing coronary angioplasty. *J Pharmacol Exp Ther* **307**(3):969–976 (2003).

33. C. Ross, K. M. Clemmesen, M. Svenson, P. S. Sorensen, N. Koch-Henriksen, G. L. Skovgaard, and K. Bendtzen, Immunogenicity of interferon-beta in multiple sclerosis patients: influence of preparation, dosage, dose frequency, and route of administration, Danish Multiple Sclerosis Study Group. *Ann Neurol* **48**:706–712 (2000).

34. J. G. Gribben, S. Devereux, N. S. Thomas, et al., Development of antibodies to unprotected glycosylation sites on recombinant human GM-CSF. *Lancet* **335**:434–437 (1990).

35. M. J. Keating, R. Holmes, S. Lerner, and D. H. Ho, L-Asparaginase and PEG asparaginase—past, present, and future. *Leuk Lymphoma* **10**(Suppl):153–157 (1993).

36. G. E. Linthorst, C. E. Hollak, W. E. Donker-Koopman, A. Strijland, and J. M. Aerts, Enzyme therapy for Fabry disease: neutralizing antibodies toward agalsidase alpha and beta. *Kidney Int* **66**(4):1589–1595 (2004).

37. J. P. Siegel, Issues in therapeutics research. *CBER Centennial Presentation*, September 23–24, 2002.

38. H. Schellekens, Immunogenicity of therapeutic proteins. *Nephrol Dial Transplant* **18**:1257–1259 (2003).

39. N. Casadevall, J. Nataf, B. Viron, A. Kolta, J. Kiladjian, P. Martin-Dupont, P. Michaud, T. Papo, V. Ugo, I. Teyssandier, B. Varet, and P. Mayeux, Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N Engl J Med* **346**(7):469–475 (2002).

40. F. W. R. Brambell, The transmission of immunity from mother to young and the catabolism of immunoglobins. *Lancet* **1966**:1089 (1966).

41. R. P. Junghans, Finally! The Brambell receptor (FcRB): mediator of transmission of immunity and protection from catabolism for IgG. *Immunol Res* **16**:29–57 (1997).

42. V. Ghetie and E. S. Ward, FcRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunol Today* **18**:592–598 (1997).

43. R. P. Junghans and C. L. Anderson, The protection receptor for IgG catabolism is the b2-microglobulin-containing neonatal intestinal transport receptor. *Proc Natl Acad Sci USA* **96**:5512–5516 (1996).

44. A. Supersaxo, W. R. Hein, and H. Steffen, Effect of molecular weight on the lymphatic absorption of water-soluble compounds following subcutaneous administration. *Pharm Res* **7**(2):167–169 (1990).

45. C. J. Porter and W. N. Charman, Transport and absorption of drugs via the lymphatic system. *Adv Drug Deliv Rev* **50**(1–2):1–2 (2001).

46. N. Hayashi, H. Kinoshita, E. Yukawa, and S. Higuchi, Pharmacokinetic and pharmacodynamic analysis of subcutaneous recombinant human granulocyte colony stimulating factor (lenograstim) administration. *J Clin Pharmacol* **39**(6):583–592 (1999).

47. C. J. Porter and S. A. Charman, Lymphatic transport of proteins after subcutaneous administration. *J Pharm Sci* **89**(3):297–310 (2000).

48. U. B. Kompella and V. H. L. Lee, Pharmacokinetics of peptide and protein drugs, in *Peptide and Protein Drug Delivery*, V. H. L. Lee (Ed.). Marcel Dekker, New York, 1991, pp. 391–484.

49. P. S. Banerjee, E. A. Hosny, and J. R. Robinson, Parenteral delivery of peptide and protein drugs, in *Peptide and Protein Drug Delivery*, V. H. L. Lee (Ed.). Marcel Dekker, New York, 1991, pp. 487–543.

50. G. Karanikas, H. Ulrich-Pur, A. Becherer, K. Wiesner, R. Dudczak, M. Raderer, and K. Kletter, Uptake of indium-111-labeled human polyclonal immunoglobulin G in pancreatic cancer: in vivo and in vitro studies. *Oncol Rep* **9**(2):353–357 (2002).

51. Y. S. Lin, C. Nguyen, J. L. Mendoza, E. Escandon, D. Fei, Y. G. Meng, and N. B. Modi, Preclinical pharmacokinetics, interspecies scaling, and tissue distribution of a humanized monoclonal antibody against vascular endothelial growth factor. *J Pharmacol Exp Ther* **288**(1):371–378 (1999).

52. *Physician's Desk Reference*, 58th ed. Thomson Healthcare, 2004.

53. B. Wang, T. M. Ludden, E. N. Cheung, G. G. Schwab, and L. K. Roskos, Population pharmacokinetic–pharmacodynamic modeling of filgrastim (r-metHuG-CSF) in healthy volunteers. *J Pharmacokinet Pharmacodyn* **4**:321–342 (2001).

54. R. Ramakrishnan, W. K. Cheung, M. C. Wacholtz, N. Minton, and W. J. Jusko, Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after single and multiple doses in healthy volunteers. *J Clin Pharmacol* **44**(9):991–1002 (2004).

55. K. A. Nieforth, R. Nadeau, I. H. Patel, and D. Mould, Use of an indirect pharmaco-dynamic stimulation model of MX protein induction to compare in vivo activity of interferon alfa-2a and a polyethylene glycol-modified derivative in healthy subjects. *Clin Pharmacol Ther* **59**(6):636–646 (1996).

56. C. Kloft, E. U. Grafe, P. Tanswell, A. M. Scott, R. Hofheinz, A. Amelsberg, and M. O. Karlsson, Population pharmacokinetics of sibrotuzumab, a novel therapeutic monoclo-nal antibody in cancer subjects. *Invest New Drugs* **22**:39–52 (2004).

57. I. F. Khouri, M. Albitar, R. M. Saliba, C. Ippoliti, Y. C. Ma, M. J. Keating, and R. E. Camplin, Low-dose alemtuzumab (Campath) in myeloablative allogeneic stem cell transplantation for CD52-positive malignancies: decreased incidence of acute graft-versus-host-disease with unique pharmacokinetics. *Bone Marrow Transplant* **33**:833–837 (2004).

58. D. S. Fefferman, P. J. Lodhavia, M. Alsahli, K. R. Falchuk, M. A. Peppercorn, S. A. Shah, and R. J. Farrell, Smoking and immunomodulators do not influence the response or duration of response to infliximab in Crohn's disease. *Inflamm Bowel Dis* **10**(4):346–351 (2004).

59. M. Koch, G. Niemeyer, I. Patel, S. Light, and B. Nashan, Pharmacokinetics, pharma-codynamics, and immunodynamics of daclizumab in a two-dose regimen in liver trans-plantation. *Transplantation* **73**(10):1640–1646 (2002).

60. J. Duconge, E. Fernandez-Sanchez, and D. Alvarez, Interspecies scaling of the mono-clonal anti-EGF receptor ior EGF/r3 antibody disposition using allometric paradigm: is it really suitable? *Biopharm Drug Dispos* **25**(4):177–186 (2004).

61. K. Ruxrungtham, M. Boyd, S. E. Bellibas, X. Zhang, A. Dorr, S. Kolis, T. Kinchelow, N. Buss, and I. H. Patel, Lack of interaction between enfuvirtide and ritonavir or ritonavir-boosted saquinavir in HIV-1-infected subjects. *Clin Pharmacol Ther* (2003).

62. X. Zhang, J. P. Lalezari, A. D. Badley, A. Dorr, S. J. Kolis, T. Kinchelow, and I. H. Patel, Assessment of drug–drug interaction potential of enfuvirtide in human immunodeficiency virus type 1-infected patients. *Clin Pharmacol Ther* **73**(2):81 (2003).

63. M. D. Pescovitz, G. Bumgardner, R. S. Gaston, R. L. Kirkman, S. Light, I. H. Patel, K. Nieforth, and F. Vincenti, Pharmacokinetics of daclizumab and mycopheno-late mofetil with cyclosporine and steroids in renal transplantation. *Clin Transplant* **17**(6):511–517 (2003).

64. S. Ropero, J. A. Menendez, A. Vazquez-Martin, S. Montero, H. Cortes-Funes, and R. Colomer, Trastuzumab plus tamoxifen: anti-proliferative and molecular interactions in breast carcinoma. *Breast Cancer Res Treat* **86**(2):125–137 (2004).

65. M. D. Pegram, G. E. Konecny, C. O'Callaghan, M. Beryt, R. Pietras, and D. J. Slamon, Rational combinations of trastuzumab with chemotherapeutic drugs used in the treat-ment of breast cancer. *J Natl Cancer Inst* **96**(10):739–749 (2004).

66. G. M. Meno-Tetang, Y. Y. Hon, S. Van Wart, and W. J. Jusko, Pharmacokinetic and pharmacodynamic interactions between dehydroepiandrosterone and prednisolone in the rat. *Drug Metabol Drug Interact* **15**(1):51–70 (1999).

67. M. H. Magee, R. A. Blum, C. D. Lates, and W. J. Jusko, Prednisolone pharmacokinetics and pharmacodynamics in relation to sex and race. *J Clin Pharmacol* **41**(11):1180–1194 (2001).

68. R. B. Sewell, N. D. Yeomans, D. J. Morgan, and R. A. Smallwood, Hepatic Kupffer cell function: the efficiency of uptake and intracellular degradation of ^{14}C labeled mito-chondria is reduced in aged rats. *Mech Ageing Dev* **73**(3):157–168 (1994).

69. A. Pascher, J. Pratschke, P. Neuhaus, and S. G. Tullius, Modifications of immune regulations with increasing donor and recipient age. *Ann Transplant* **9**(1):72–73 (2004).

70. F. E. Fox, D. Mauro, S. Bai, R. Raymond, C. Farrell, H. Pentikis, A. Nolting, M. Birkhofer, and M. Needle, A population pharmacokinetic (PPK) analysis of the anti-EGFr specific IgG1 monoclonal antibody cetuximab. ASCO 2004 Abstract 290.

71. N. L. Dayneka, V. Garg, and W. J. Jusko, Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* **21**(4):457–478 (1993).

72. G. Movin-Osswald and M. Hammarlund-Udenaes, Prolactin release after remoxipride by an integrated pharmacokinetic–pharmacodynamic model with intra- and interindividual aspects. *J Pharmacol Exp Ther* **274**(2):921–927 (1995).

73. A. Sharma, W. F. Ebling, and W. J. Jusko, Precursor-dependent indirect pharmacodynamic response model for tolerance and rebound phenomena. *J Pharm Sci* **87**(12):1577–1584 (1998).

74. K. Sandstrom, L. Tuason, C. Redovan, P. Rayman, R. Tubbs, D. Resta, P. Elson, and J. Finke, Phase I trial of subcutaneous interleukin 3 in patients with refractory malignancy: hematological, immunological, and pharmacodynamic findings. *Clin Cancer Res* **2**(2):347–357 (1996).

75. S. C. Piscitelli, A. Forrest, S. Vogel, D. Chaitt, J. Metcalf, R. Stevens, M. Baseler, R. T. Davey, and J. A. Kovacs, Pharmacokinetic modeling of recombinant interleukin-2 in patients with human immunodeficiency virus infection. *Clin Pharmacol Ther* **64**(5):492–498 (1998).

76. A. Isaccs, J. Lindenmann, and R. C. Valentine, Virus interference. II. Some properties of interferon. *Proc R Soc Lond B Biol Sci* **147**(927):268–273 (1957).

77. C. E. Samuel, Antiviral actions of interferons. *Clin Microbiol Rev*. **14**(4):778–809 (2001).

78. U. Muller, U. Steinhoff, L. F. Reis, S. Hemmi, J. Pavlovic, R. M. Zinkernagel, and M. Aguet, Functional role of type I and type II interferons in antiviral defense. *Science* **264**(5167):1918–1921 (1994).

79. R. T. Dorr, Interferon-alpha in malignant and viral diseases. A review. *Drugs* **45**(2):177–211 (1993).

80. K. M. Zsebo, A. M. Cohen, D. C. Murdock, T. C. Boone, H. Inoue, V. R. Chazin, D. Hines, and L. M. Souze, Recombinant human granulocyte colony-stimulating factor: molecular and biological characterization. *Immunobiology* **172**:175–184 (1986).

81. K. Welte, M. A. Bonilla, A. P. Gillio, T. C. Boone, G. K. Potter, J. L. Gabrilove, M. A. Moore, R. J. O'Reilly, and L. M. Souza, Recombinant human G-CSF: effects on hematopoiesis in normal and cyclophosphamide treated primates. *J Exp Med* **165**:941–948 (1987).

82. U. Duhrsen, J. L. Villeval, J. Boyd, G. Kannourakis, G. Morstyn, and D. Metcalf, Effects of recombinant human granulocyte colony-stimulating factor on hematopoietic progenitor cells in cancer patients. *Blood* **72**:2074–2081 (1988).

83. L. M. Souza, T. C. Boone, J. Gabrilove, et al., Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. *Science* **232**:61–65 (1986).

84. A. C. Kotto-Kome, S. E. Fox, W. Lu, B. B. Yang, R. D. Christensen, and D. A. Calhoun, Evidence that the granulocyte colony-stimulating factor (G-CSF) receptor plays a role in the pharmacokinetics of G-CSF and PegG-CSF using a G-CSF-R KO model. *Pharmacol Res* **50**(1):55–58 (2004).

85. T. Kuwabara, S. Kobayashi, and T. Sugiyama, Pharmacokinetics and pharmacodynamics of a recombinant human granulocyte colony-stimulating factor. *Drug Metab Rev* **28**(4):625–658 (1996).

86. U. Wahlby, A. H. Thomson, P. A. Milligan, and M. O. Karlsson, Models for time-varying covariates in population pharmacokinetic-pharmacodynamic analysis. *Br J Clin Pharmacol* **58**(4):367–377 (2004).

87. B. Jilma, N. Hergovich, M. Homoncik, P. Jilma-Stohlawetz, C. Kreuzer, H. G. Eichler, M. Zellner, and J. Pugin, Granulocyte colony-stimulating factor (G-CSF) downregulates its receptor (CD114) on neutrophils and induces gelatinase B release in humans. *Br J Haematol* **111**:314–320 (2000).

88. L. E. Friberg, A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson, Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol* **20**(24):4713–4721 (2002).

89. S. Masuda, M. Okano, K. Yamagishi, M. Nagao, M. Ueda, and R. Sasaki, A novel site of erythropoietin production, oxygen-dependent production in cultured rat astrocytes. *J Biol Chem* **269**:19488–19493 (1994).

90. T. Lappin, The cellular biology of erythropoietin receptors. *Oncologist* **8**(Suppl 1):15–18 (2003).

91. F. Pajonk, A. Weil, A. Sommer, R. Suwinski, and M. Henke, The erythropoietin-receptor pathway modulates survival of cancer cells. *Oncogene* **23**(55):8987–8991 (2004).

92. G. Acs, X. Xu, C. Chu, P. Acs, and A. Verma, Prognostic significance of erythropoietin expression in human endometrial carcinoma. *Cancer* **100**(11):2376–2386 (2004).

93. V. Brower, Erythropoietin may impair, not improve, cancer survival. *Nat Med* **9**(12):1439 (2003).

94. W. Jelkmann, The enigma of the metabolic fate of circulating erythropoietin (EPO) in view of the pharmacokinetics of the recombinant drugs rhEPO and NESP. *Eur J Haematol* **69**:265–274 (2002).

95. P. Veng-Pedersen, S. Chapel, N. H. Al-Huniti, R. L. Schmidt, E. M. Sedars, R. J. Hohl, and J. A. Widness, Pharmacokinetic tracer kinetics analysis of changes in erythropoietin receptor population in phlebotomy-induced anemia and bone marrow ablation. *Biopharm Drug Dispos* **25**(4):149–156 (2003).

96. R. Ramakrishnan, W. K. Cheung, M. C. Wacholtz, N. Minton, and W. J. Jusko, Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after single and multiple doses in healthy volunteers. *J Clin Pharmacol* **44**(9):991–1002 (2004).

97. L. S. Valberg, R. T. Card, E. J. Paulson, and J. Szivek, The metal composition of erythrocytes in different species and its relationship to the lifespan on the cells in the circulation. *Comp Biochem Physiol* **15**(3):347–359 (1965).

98. W. Krzyzanski, R. Ramakrishnan, and W. J. Jusko, Basic pharmacodynamic models for agents that alter production of natural cells. *J Pharmacokinet Biopharm* **27**(5):467–489 (1999).

99. V. Renard, A. Cambiaggi, F. Vely, M. Blery, L. Olcese, S. Olivero, M. Bouchet, and E. Vivier, Transduction of cytotoxic signals in natural killer cells: a general model of fine tuning between activatory and inhibitory pathways in lymphocytes. *Immunol Rev* **155**:205–221 (1997).

100. M. Zhang, Z. Zhang, K. Garmestani, C. K. Goldman, J. V. Ravetch, M. W. Brechbiel, J. A. Carrasquillo, and T. A. Waldmann, Activating Fc receptors are required for antitumor efficacy of the antibodies directed toward CD25 in a murine model of adult T-cell leukemia. *Cancer Res* **64**(16):5825–5829 (2004).

101. M. J. Glennie and P. W. Johnson, Clinical trials of antibody therapy. *Immunol Today* **21**(8):403–410 (2000).

102. H. Waldmann and S. Cobbold, How do monoclonal antibodies induce tolerance? A role for infectious tolerance. *Annu Rev Immunol* **16**:619–644 (1998).

103. A. J. Boeckmann, S. L. Beal, and L. B. Sheiner, *NONMEM Users Guide*, *Parts I–VIII*, University of California, San Francisco, 1998.

Analysis of Quantic Pharmacokinetic Study: Robust Estimation of Tissue-to-Plasma Ratio

HUI-MAY CHU and ENE I. ETTE

42.1 INTRODUCTION

Preclinical pharmacokinetic (PK) studies provide information useful for supporting efficacy and safety evaluation studies in animals, preclinical and clinical study designs, dosing regimen development, and interpretation of toxicity data. These studies provide PK data that may be useful in dose escalation in healthy volunteers and patients. Toxicokinetics is a major component of toxicology studies. It enables the investigation of the relationship between drug dose and measured concentration, primarily the establishment of the dose proportionality and linearity or nonlinearity in pharmacokinetics.

Toxicokinetic and PK research studies are characterized by some uncertainty regarding the process studied and significant variation in the concentration measurements obtained. Variability in PK parameters among homogeneous strains of small laboratory animals has been reported to be between 30% and 50% in some cases (1, 2). In addition to the inherent variability of the biological system, there is the uncertainty associated with the assay and process noise.

The number of samples that can be obtained per subject is limited to one sample per subject (especially when destructive sampling is implemented) in most rodent toxicokinetic studies. The fact is that, for small laboratory animals, the periods between successive sampling times are simply not long enough to allow sufficient recovery. A major disadvantage of this sampling scheme is that intraindividual concentration–time profiles are unavailable. This poses a data analysis challenge because the one sample per subject data constitutes the extreme case of sparsely sampled PK data, hence extremely sparse data, with independent observations over time. The situation is complicated when tissue sampling (e.g., in tissue distribution studies) is involved, and the ratio of tissue to plasma concentrations is the object of

the investigation. Equally, only one tissue sample/subject is obtained in such studies because the animal is usually sacrificed.

A solution for analyzing extremely sparse data is to use the nonlinear mixed effects modeling approach. This has been elegantly addressed in the literature (3–6). However, it is common practice that to compute a PK parameter such as area under the concentration–time curve (AUC) some form of data pooling is used. Thus, to compute noncompartmental AUC, actually "composite" AUC, data are averaged at each time point and the parameter is estimated using the trapezoidal rule. This composite approach to the estimation of AUC has problems associated with it. The AUCs and other mean PK parameters are estimated with no measures of uncertainty associated with them. Other measures characterizing the distribution of the parameters are, in general, difficult to obtain. Some ad hoc solutions have been proposed for the estimation of the average (or "typical") AUC in a population of extremely sparsely sampled subjects (7–11). Pai et al. (12) proposed the use of the bootstrap resampling technique for the estimation of AUC from sparsely sampled populations in toxicology studies.

42.2   ESTIMATION OF TISSUE-TO-PLASMA RATIO

The challenge in the analysis of quantic (one sample/subject) data is further complicated when tissue sampling (e.g., in tissue distribution studies) is involved, and the ratio of tissue to plasma concentrations is the object of the investigation. Equally, only one tissue sample/subject is obtained in such studies because the animal is usually killed.

Tissue-to-plasma ratio is commonly determined from the ratio of average concentrations at specified time points. It is not uncommon, in practice, for the ratios to be calculated at selected time points corresponding to peak and trough concentrations, and the variations in the ratios are usually very large. This finding could be attributed in part to the variations in the concentrations and a lack of accounting for the correlation in observations from the biological matrices sampled from each subject.

Occasionally, tissue-to-plasma ratio is calculated using area under the concentration curves (AUCs) calculated from mean profiles using the noncompartmental approach. These "composite" AUCs are usually computed from data that are averaged at each time point (naive data averaging approach) using the trapezoidal rule. The tissue-to-plasma ratios computed using either average concentrations at specified time points or composite AUC values are usually reported without regard to the correlation structure in the data, and no measures of dispersion and uncertainty associated with them.

In the sections that follow a data set from a quantic PK study is assumed and various approaches used to estimate tissue-to-plasma ratio are discussed as they are applied to the data. Thus, the approaches used in the estimation of tissue-to-plasma ratio are presented with their advantages and disadvantages; the importance of examining convergence with resampling algorithms is discussed, as well as the impact of outliers on the performances of the ratio estimation approaches. This is followed with an overall thought on the estimation of tissue-to-plasma ratio with a recommendation on a preferred approach (a nonparametric random sampling

approach) for the robust estimation of robust tissue-to-plasma ratio in a drug development setting.

42.2.1 Data Set

Since the objective of this chapter is the development of a methodology for estimating robust tissue-to-plasma ratio in a drug development setting, such details as are necessary for understanding the proposed methodology are presented. It is important to note that in drug development, pragmatism, efficiency, and effectiveness are major considerations.

A toxicokinetic study was performed to determine the tissue-to-plasma ratio of a drug in development. An oral dose of a drug in development was administered to 18 rats, and each animal was killed at one of six specified time points: 0.5, 1, 2, 4, 6, and 8 hours. Therefore, each animal had only 1 pair of concentrations, 1 each from plasma and tissue, respectively. Table 42.1 shows the data set from the study used in our investigation. The effect of correlation structure in the data set is also of interest. Thus, we investigated the effect of maintaining or breaking the relationship between tissue and plasma concentrations within the same animal, using both paired and unpaired tissue and plasma data to evaluate the effect on the robustness of estimation of tissue-to-plasma ratio.

42.2.2 Approaches for Estimating Tissue-to-Plasma Ratio

We have taken a very practical approach in addressing the computation of tissue-to-plasma ratio in a drug development setting. Thus, approaches that are commonly used in practice (i.e., the naive data averaging and ratios of concentrations by time point approaches) for computing tissue-to-plasma ratio were employed in this investigation and compared with our proposed methodology—the random sampling approach. Since our approach is a sampling-based approach, we have included a comparison of the performance of our approach with another sampling-based approach reported in the literature, the PpbB (14), in the estimation of tissue-to-plasma ratio. First, the naive data averaging approach is discussed, followed by a discussion of the random sampling approach, and then the PpbB approach.

TABLE 42.1   A Sample Data Set from an Oral Toxicokinetic Study[a]

Biological Matrix	Time Point (hour)					
	0.5	1	2	4	6	8
Plasma	0.18	0.13	0.12	0.04	0.00	0.01
Tissue	9.05	1.76	1.26	0.18	0.02	0.42
Plasma	0.18	0.14	0.11	0.03	0.05	0.00
Tissue	5.24	1.65	1.67	0.64	0.28	0.07
Plasma	0.17	0.18	0.05	0.02	0.02	0.00
Tissue	2.92	4.18	0.74	0.19	0.00	0.10

[a]Each cell represents a pair of values from one animal; there are 18 animals in total.

All methods were implemented in the statistical software, S-Plus Version 6.02 (Insightful, Seattle, WA).

42.2.3 Naive Data Averaging Approach

The approach involves computing the average value of the data for each sampling time k:

$$\overline{C}_k = \frac{1}{I}\sum_{i=1}^{I} C_{ik}$$

for $i = 1, \ldots, I$, where I is the standard number of individual subject data at time point k. The averaging of data across subjects is a common practice owing to the assumption that all concentrations at each time point have been measured under identical conditions.

Thus, tissue-to-plasma ratio is estimated independently for each time point using the averaged concentration at each time point. Alternatively, the noncompartmental AUC, actually the composite AUC, can be estimated using the trapezoidal rule. From this point, the use of the term "naive data averaging approach" will be reserved for estimation of AUC. The term "unpaired independent time points approach" will be reserved for use in cases where tissue-to-plasma ratio is calculated at each time point using a measure of central tendency (mean or median) of the measured concentrations without regard to the correlation structure in the observations. The term "paired independent time points" approach will be used when the pairing of observations is taken into account in the calculation of the tissue-to-plasma concentration ratio at each time point.

42.2.4 Traditional Naive Data Averaging Approach Incorporating Independent Time Points Approaches

The results of tissue-to-plasma ratio values obtained with three approaches: (a) unpaired independent time points approach, (b) paired independent time points approach, and (c) naive data averaging approach are presented herewith. Table 42.2 illustrates ratios obtained across time points by calculating the mean and median for each time point independently for tissue and plasma. There is no measure of variability around each time point, as expected. When zero was returned for plasma concentration (e.g., 6 and 8 hour time points in Table 42.2) because the levels were not quantifiable or below the limit of quantification, the zero divisor of the ratio yielded the result #DIV/0. Such an outcome cannot be interpreted and is usually discarded in the presentation of results with the independent time points approach. Table 42.3 contains summary statistics of paired tissue-to-plasma ratios obtained with the paired independent time points approach. The tissue-to-plasma ratios by time point can vary from 0 to 50.3 across different time points, as shown in the last row of Table 42.3. Table 42.4 contains the tissue-to-plasma AUC ratio (TPAR) derived by calculating the mean AUC values using the naive data averaging approach across time points for both tissue and plasma without regard to the correlation structure in the data. As expected, there is also no measure of variability around TPAR obtained with the naive data averaging method. The code for implementing the naive data averaging approach is in Appendix 42.1.

TABLE 42.2   Tissue-to-Plasma Ratios Calculated Using the Unpaired Independent Time Points Approach[a]

Tissue/Plasma Ratio	Time Point (hour)					
	0.5	1	2	4	6	8
Median	30	14	14	10	`#DIV/0`	`#DIV/0`
Mean	32	16	13	12	`#DIV/0`	`#DIV/0`

[a] `#DIV/0` indicates the denominator (plasma concentration) of the ratio is 0 (or below the quantifiable limit—BQL).

TABLE 42.3   Tissue-to-Plasma Ratios Calculated Using the Paired Independent Time Points Approach

Time (hours)	Ratios[a]				
	Minimum	Q1	Mean	Q3	Maximum
0.5	17	23.4	32.3	40	50.3
1	12.2	13.2	16.4	18.5	23
2	10.3	12	13.4	14.9	16
4	5.2	7.4	11.8	15.2	20.7
6	0	1.5	2.9	4.4	5.9
8	32.6	32.6	32.6	32.6	32.6
All time points	0	10	17.4	21.8	50.3

[a] Q1 indicates first quartile and Q3 the third quartile.

TABLE 42.4   Tissue-to-Plasma Ratios Calculated from AUC Values Obtained Via the Naive Data Averaging Approach

Central Tendency	Plasma AUC[a]	Tissue AUC	Tissue-to-Plasma AUC Ratio
Mean	0.45	7.62	17.05

[a] AUC indicates area under the curve.

42.2.5   Random Sampling Approach

The random sampling (RS) approach was recently proposed by Chu and Ette (13). To implement the approach, the population sampling pool is first generated, and it comprises a large set of individual pharmacokinetic (PK) profiles based on the empirical data by resampling with replacement. This potential population pool contains M_1 copies of PK profiles for each subject to ensure equal opportunity for each subject to be resampled for the next step. Next, M_2 copies of the virtual study are drawn from the population pool, and then any function of interest is computed from the virtual study level. Figure 42.1 is a schematic chart illustrating the RS approach. The RS algorithm, therefore, is defined in two phases.

FIGURE 42.1 Schematic chart for random sampling approach.

Phase 1: Setting Up the Population Sampling Pool by Generating Individual Subject Sampling Pools

Phase 1 is done by constructing the individual level sampling pool (i.e., the concentration values for the ith subject at rth replicate resampling (C^*_{ir})). The steps to do this are as follows:

Step 1. For the ith subject with datum observed at time point w, randomly resample M_1 times with replacement from the available values independently at each time point that the subject had no observation. For a subject that has wth time point observation, for example, the concentration values are to be resampled (i.e., plasma and tissue concentrations) at other k time points C^*_{ikr}, (where, $k = 1, \ldots, K$, but $k \neq w$, and $r = 1, \ldots, M_1$) to create a "complete profile" encompassing all sample points, including the observed C_{iwr} and the resampled vector C^*_{ikr}. More specifically, $C^*_{i..}$, M_1 replicates of "complete profiles" for the ith subject, can be expressed as the following matrix:

$$C^*_{i..} = \begin{bmatrix} c^*_{i11} & \ldots & c_{iw1} & \ldots & c^*_{iK1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c^*_{i1M_1} & \ldots & c_{iwM_1} & \ldots & c^*_{iKM_1} \end{bmatrix}$$

Each row represents one profile encompassing all sample/time points. Each column is M_1 copy of the same time point.

Step 2. Repeat Step 1 of Phase 1 to construct the individual profile pool for each subject.

Step 3. Calculate functions of interest from each profile (e.g., AUC, C_{max}).

The population sampling pool of complete profiles is now ready to be sampled for the next phase of virtual study resampling.

Phase 2: Generation of Samples at the Study Level

Step 1. Draw M_2 copies of size N (where N is the sample size, total number of animals, in the real study) of functions of interest from the population sampling pool obtained from Phase 1.

Step 2. Calculate the summary statistics (i.e., quantiles, mean, and median) of the function of interest from each virtual study obtained from Step 1 of Phase 2.

Step 3. Derive the summary statistics of required parameters across virtual studies with their associated standard deviations.

The S-Plus code used in implementing the RS approach is in Appendix 42.2.

42.2.6   Pseudoprofile-Based Bootstrap

The PpbB approach (14) generates estimates of both the distributions of the raw data and the corresponding measures of variability. The term "pseudoprofile" was applied to the information obtained when one sample is obtained per animal but several animals are sampled at each of several times postdose.

Bootstrap resampling is performed twice within the PpbB approach to generate PK pseudoprofiles from which the function of interest is estimated. More specifically, the following scheme is adopted for the b_1th replicate at each time point:

Step 1. Resample with replacement at one concentration, denoted as $C^*_{b1}(t_k)$, at time t_k for $k = 1$ to K from the respective concentration vectors and keep K concentrations. $c^*_{b1}(t_k)$, $k = 1, \ldots, K$.

Step 2. Construct a pseudoprofile, that is, $c^*_{b1} = \{c^*_{b1}(t_1), c^*_{b1}(t_2), \ldots, c^*_{b1}(t_{k-1}), c^*_{b1}(t_k)\}$.

Step 3. Repeat Steps 1 and 2 B_1 times to generate a PK pseudoprofile pool $\hat{F}^*$, an estimate of the distribution F.

Step 4. Calculate a function of interest from each pseudoprofile (i.e., AUC, $C_{\max}$).

Step 5. Perform B_2 times bootstrap resampling with replacement from this empirical distribution $\hat{F}^*$ with sample size $\bar{n}$ each, where $\bar{n}$ is the average number of concentration replicates, and the corresponding parameter for each $b_2 = 1, \ldots, B_2$ is estimated.

Step 6. Calculate the bootstrap estimates of the mean parameter and its standard deviation.

The S-Plus code used in implementing the PpbB approach is in Appendix 42.3.

Given the limitations associated with the naive data averaging approaches in estimating the tissue-to-plasma ratio, the RS approach is compared with the PpbB approach in subsequent sections. However, occasional references are made to the naive data averaging and independent time points approaches because of their use in common practice.

42.3   COMPARISON OF PpbB AND RS APPROACHES

42.3.1   Paired Versus Unpaired Data

Figure 42.2 illustrates the results obtained with the RS approach using paired and unpaired data at different replication levels (i.e., M_1 equals 10, 100, and 500 to build up the population pool). This was then followed by a calculation of TPAR over M_2 (i.e., 50) virtual studies (with $N = 18$ for each study). Across all three population pool levels (i.e., $M_1 = 10$, 100, and 500), paired observations consistently yielded tighter distributions than unpaired ones. Similar results were obtained with the PpbB approach.

If a drug is designed to target a particular tissue, the interest might be in having a minimal target TPAR. In that case, having knowledge of mean TPAR would not be enough. Having knowledge of the distribution of TPAR across virtual studies (i.e., replicates) in terms of the summary statistics (first quartile Q1, mean, median, third quartile Q3) becomes valuable. Thus, knowing that the TPAR is not below a certain cutoff, such as the first quartile of the TPAR distribution, would be important. To provide such an insight, we examined the distribution of TPAR within and between replicates and have provided a summary of the distribution of TPAR across virtual studies. Consequently, Figure 42.3 provides an amplification of the outcomes with the two approaches when the first and third quartiles (Q1 and Q3, respectively) for paired and unpaired data are compared. The quartiles for the unpaired data have a

FIGURE 42.2   Computation of tissue-to-plasma AUC ratios from paired and unpaired data using the random sampling approach. The line inside the box represents the median, and the box represents the limits of the middle half of the data. The range of the box, from the first quartile (Q1) to the third quartile (Q3), is called the interQuartile range (IQR). The standard span of the data is defined within the range from Q1 − 1.5IQR to Q3 + 1.5IQR. Whiskers, the dotted line, are drawn to the nearest value not beyond the range of the standard span; points beyond (outside values) are drawn individually.

FIGURE 42.3 Comparison of the performance of the RS and PpbB approaches when tissue-to-plasma AUC ratios (TPAR) are computed from paired and unpaired data. The comparison is focused on distribution of the first and third quartiles (Q1 and Q3, respectively) of TPARs.

wider spread, with the lower adjacent value of the distribution of Q1 values in the box plot extending beyond that for paired data in both RS and PpbB approaches. Disrupting the correlation structure in the data by unpairing the data yielded more variable results than when the correlation structure in the data was maintained by pairing. Thus, breaking the correlation structure between tissue and plasma observations resulted in a loss of information. Therefore, the rest of the study is focused on the paired scenario only.

A tabular comparison of the results obtained with the RS and PpbB approaches is shown in Table 42.5. In addition to the typical fashion of only describing distribution of mean of TPAR, Table 42.5 also includes distributions of quartiles of TPAR in terms of Q1 and Q3 with associated standard errors. The resampling approaches yielded comparable results when the number of replications was at least 600 with mean TPAR around 17, but the RS approach converged faster than the PpbB approach. (See Section 42.3.2 for more details.)

42.3.2 Convergence

Convergence was determined for both RS and PpbB approaches. That is, the number of replications (i.e., the number of times the sampling/resampling has to be repeated) needed for stable estimates of tissue-to-plasma AUC ratio (TPAR) to be obtained was determined for both methods. An empirical approach was used to determine convergence (13).

TABLE 42.5 Distribution of Tissue-to-Plasma AUC Ratio Parameter Estimates Obtained Using Random Sampling and Pseudoprofile-Based Bootstrap Approaches

	Random Sampling Approach[a]: Tissue-to-Plasma AUC Ratios						Pseudoprofile-Based Bootstrap[a]: Tissue-to-Plasma AUC Ratios					
	Q1		Mean		Q3		Q1		Mean		Q3	
Repl[b]	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)
5	14.31	(0.94)	16.64	(0.88)	18.64	(1.45)						
10	15.01	(0.95)	17.51	(0.85)	19.80	(1.29)	14.43	(0.42)	15.76	(0.41)	17.15	(1.10)
50	14.46	(1.00)	17.17	(0.95)	19.33	(1.48)	14.05	(0.75)	16.71	(0.83)	18.70	(1.09)
100	14.75	(0.96)	17.28	(0.82)	19.38	(1.21)	14.12	(0.93)	17.93	(0.68)	20.36	(0.95)
200	14.57	(0.99)	17.15	(0.84)	19.35	(1.26)	14.45	(1.15)	17.02	(0.85)	19.07	(1.20)
300	14.83	(1.01)	17.39	(0.90)	19.71	(1.32)	14.63	(0.94)	17.16	(0.87)	19.20	(1.34)
400	14.34	(1.12)	17.08	(0.95)	19.28	(1.46)	14.89	(0.93)	17.27	(0.80)	19.35	(1.16)
500	14.49	(0.89)	17.16	(0.84)	19.48	(1.36)	14.28	(0.85)	17.08	(0.78)	19.54	(1.29)
600	14.53	(1.08)	17.13	(0.91)	19.28	(1.17)	14.57	(0.96)	17.30	(0.80)	19.76	(1.19)
700	14.61	(1.05)	17.25	(0.97)	19.53	(1.38)	14.34	(1.09)	17.05	(0.88)	19.38	(1.33)
800	14.84	(0.99)	17.41	(0.84)	19.61	(1.20)	14.90	(1.02)	17.44	(0.91)	19.77	(1.25)
900	14.75	(1.05)	17.24	(0.89)	19.49	(1.24)	14.50	(0.94)	17.12	(0.84)	19.32	(1.34)
1000	14.43	(1.01)	17.09	(0.87)	19.41	(1.35)	14.53	(1.02)	17.18	(0.78)	19.43	(1.17)

[a] Q1 and Q3 refer to the first and third quartiles of distribution of TPAR; AUC indicates area under the curve; and Repl stands for replication.
[b] Replication is M_1 (i.e., the number of replicates for each subject in Phase 1 of the RS approach).
[c] Replication represents B_1 replicates used in the PpbB approach.

To examine the effect of the number of replications (i.e., M_1 in RS and B_1 in PpbB), a graphical presentation of percentage change (PC) of mean TPAR is shown in the middle panel of Figure 42.4 and Figure 42.5 for the RS and PpbB approaches, respectively. In addition, the PC values of Q1 and Q3 are also plotted in the left and right panels of each figure. The acceptable range for the percentage change is calculated from summary statistics/confidence intervals of PC across all replication levels considered (i.e., from M_1 with as little as 5 replications to as high as 1000 replications), and for statistics Q1, mean, and Q3. This range was determined by visual inspection of the convergence graphs with the assumption that, over the range of the replications, the PC trend should be stabilized with limited amount of fluctuations. Therefore, the percentile cutoff range was chosen using a trimming approach, and the range of percentiles 12.5 and 87.5 was found to be appropriate for both sampling approaches and across the three summary statistics. Figure 42.4 shows the convergence trend for the RS approach. For all three statistics (Q1, mean, and Q3) of interest, 100 replications are sufficient. On the other hand, the number of replications needed for the distributions of summary statistics of TPAR with the PpbB is at least 600 replications (see Figure 42.5), owing to the instability in Q1. The range for the RS approach is considerably tighter than that for the PpbB approach. In fact, the range of PC is −1.28% to 1.56% for the RS approach, and −2.26% to 5.10% for the PpbB approach. This finding indicates that there was a larger variability in TPAR estimates obtained with the PpbB approach when compared with that obtained using the RS approach. The uniqueness of the RS approach lies in the population sampling pool, which is populated by generating M_1 replications

FIGURE 42.4   Convergence trend monitoring using percentage change in summary statistics (Q1, mean, and Q3) of tissue-to-plasma AUC ratio (TPAR) estimates obtained by the random sampling approach.

FIGURE 42.5   Convergence trend monitoring using percentage change in summary statistics (Q1, mean, and Q3) of distribution of tissue-to-plasma AUC ratio (TPAR) estimates obtained by the PpbB approach.

through resampling concentration–time profiles for each subject (i.e., in this study with $N = 18$ animals, 100 replications for each animal is equivalent to total of 1800 [$= 18 \cdot 100$] distinct PK profiles in the population sampling pool). Also, M_2 copies of virtual studies are sampled from the population sampling pool to derive a distribution for any function of interest. The code for monitoring convergence is shown in Appendix 42.4.

42.3.3   Outlier Effect on Robustness

To investigate the effect of outliers on the robustness with which TPAR was estimated with naive averaging, RS, and PpbB approaches, new data sets were simulated by introducing outlier(s) into the data set. The scenarios we chose can be mapped as a grid (2×4 table) (i.e., one or two outliers produced by inflating the higher tissue concentration time points by 10%, 20%, 30%, or 40%). The higher tissue concentration time points were defined as concentrations obtained within 4 hours postdose. These concentrations were randomly chosen in each replication. The outliers were introduced in the region of the concentration–time profile (i.e., around the higher concentrations), where they were likely to produce maximum effect (see S-Plus code in the Appendix 42.5).

Figure 42.6 shows the distribution of mean TPAR obtained from simulating 50 replicates (i.e., $M_2 = 50$) of the base data set with the value of one tissue concentration value inflated to create an outlier in each replicate. The effect of one outlier can be measured by how big the distance is from the original mean TPAR value of

FIGURE 42.6 The effect of outlier on distribution of tissue-to-plasma AUC ratios (TPAR) when inflating one concentration by (A) 10%, (B) 20%, (C) 30%, or (D) 40% using the naive averaging (naive), PpbB, and RS approaches.

~17 (see Table 42.5). The naive averaging approach performed the worst of all three approaches, and PpbB had a wider spread than the RS approach. In Figure 42.6, it appears that the distribution of TPAR estimates obtained with the naive averaging approach was the tightest. It has to considered that by the very nature of the naive averaging approach variability has been eliminated, hence the results. When the scenario for two outliers was considered, Figure 42.7 illustrates the effect when two tissue concentration values were randomly selected to create outliers in each replicate by calculating the PC from mean TPAR of 17 across the three methods, given the four levels of outlier perturbation (i.e., 10%, 20%, 30%, or 40% increase in concentration values). Clearly, the RS approach provides results that are more robust than the other two. The bias in the estimation of TPAR is more prominent with the PpbB and naive averaging approaches than with the RS approach (Figure 42.7).

42.4   OVERALL ASSESSMENT OF TISSUE-TO-PLASMA RATIO ESTIMATION

A nonparametric random sampling approach proposed by Chu and Ette (13) for the estimation of robust TPAR was compared with the PpbB and naive averaging approaches. Also, the estimation of tissue-to-plasma ratio using the independent time points approach was examined. It is obvious from Tables 42.2 and 42.3 that estimating tissue-to-plasma ratio independently at various times is a very

FIGURE 42.7   The effect of outlier based on the percent increase distribution of tissue-to-plasma AUC ratios (TPAR) when inflating two concentrations by 10%, 20%, 30%, or 40% using the naive averaging (left panel), PpbB (middle panel), and RS (right panel) approaches.

unreliable method, since various ratios are obtained at various time points and it is unclear which of the ratios to choose. Also, it is impossible to compute ratios when samples from a particular biological matrix are below the limit of quantification or are unquantifiable. The independent time points approach for calculating tissue-to-plasma ratio should, therefore, be avoided. Although the naive data averaging approach for computing AUC provides a single AUC value for drug exposure in each of the two matrices and consequently a single value of TPAR, the correlation in the data structure is unaccounted for and there is no measure of variability or uncertainty around the estimates. With this method, when concentrations are below the limit of quantification, they are usually ignored in the calculation of the mean concentration at the particular time point. The mean concentration is calculated only with available data. Thus, the mean profile obtained in such a situation does not represent the actual mean profile since mean concentrations at each time point are not calculated from an equal number of time points. These drawbacks not withstanding, the approach is better than the independent time points approach. However, both approaches are inferior to the resampling approaches—PpbB and RS. Breaking the correlation structure between tissue and plasma observations results in a loss of information when using any of the resampling approaches. There-fore, it is important to maintain the correlation structure in paired data sets used in estimating TPAR. By doing this, variability in the calculated TPAR is minimized.

Although there are similarities in the TPAR estimates produced by the PpbB and RS approaches, the latter converges faster than the former. Convergence is achieved with only 100 replications (i.e., $M_1 = 100$) per subject with the RS approach,

while at least 600 bootstrap (i.e., $B_1 = 600$) replications are required for the PpbB approach. In general, 100 replications are adequate in the first phase of the RS approach for robust estimation of TPAR. Also, the acceptable range for TPAR estimates is narrower for the RS approach (Figure 42.4) when compared with the PpbB approach (Figure 42.5). Thus, the PpbB approach requires a larger number of replications to yield robust estimates. The difference lies in the two-phase—population and study level—sampling of the RS approach. The tightness of the distribution of estimates obtained with the RS approach can be attributed to the creation of the representative population sample pool for subsequent study level sampling of parameters of interest. This is a unique feature of the RS approach. Also, the estimation of TPAR by the RS approach is not affected by missing data or imbalance in the number of concentrations at each time point over the sampling duration. Individual PK profiles are generated by sampling from available data at each time point across time points. Similarly, the PpbB approach is not affected by missing data or data imbalance.

When the effect of outliers on robustness was investigated, the naive data averaging approach performed the worst, while the RS approach performed the best. The edge that the RS approach has over the PpbB approach is, again, owing to the two-phase nature of implementation of the methodology. The robustness of the RS approach lies in the creation of the population sample pool before the study level sampling for the estimation of TPAR. Also, the greater bias obtained with the PpbB approach when compared with the RS approach is probably due to the fact that mean parameter estimates obtained from bootstrap replicates may be influenced by data in the tails of the distribution (15).

42.5 SUMMARY

Traditional approaches used in the estimation of TPAR have been compared with the PpbB approach and the recently proposed RS approach. The traditional approaches—independent time points and naive data averaging approaches—are inferior to the sampling/resampling approaches. The RS approach performed better than the PpbB approach because of its unique algorithm. Also, fewer replications are required for robust estimation of TPAR. The computer intensive methods provide estimates of TPAR with measures of dispersion and uncertainty. The RS approach is the method of choice for obtaining robust estimates of TPAR, when analyzing extremely sparsely sampled data.

REFERENCES

1. L. T. Lindstrom and D. S. Birkes, Estimation of population pharmacokinetic parameters using destructively obtained data: a simulation study of the one compartment open model. *Drug Metab Rev* **15**:195–264 (1984).
2. R. D. McArthur, Parameter estimation in a two compartment population pharmacokinetic model with destructive sampling. *Math Biosci* 91–157 (1988).
3. E. I. Ette EI, A. W. Kelman, C. W. Howie, and B. Whiting, Analysis of animal pharmacokinetic data: performance of the one point per animal design. *J Pharmacokinet Biopharm* **23**(6):551–566 (1995).

4. E. I. Ette EI, A. W. Kelman, C. W. Howie, and B. Whiting, Efficient experimental design and estimation of population pharmacokinetic parameters in preclinical animal studies. *Pharm Res* **12**(5):729–737 (1995).

5. E. I. Ette, A. W. Kelman, C. W. Howie, and B. Whiting, Inter-animal variability and parameter estimation in preclinical animal pharmacokinetic studies. *Clin Res Regul Affairs* **11**(2):121–139 (1994).

6. E. I. Ette, A. W. Kelman, C. W. Howie, and B. Whiting, Interpretation of simulation studies for efficient estimation of population pharmacokinetic parameters. *Ann Pharmacother* **27**:1034–1039 (1993); and Correction **27**:1548 (1993).

7. J. V. Bree, J. Nedelman, J.-L. Steimer, F. Tse, W. Robinson, and W. Niederberger, Application of sparse sampling approaches in rodent toxicokinetics: a prospective view. *Drug Info J* **28**:263–279 (1994).

8. C. D. Jones, H. Sun, and E. I. Ette, Designing cross-sectional pharmacokinetic studies: implications for pediatric and animal studies. *Clin Res Regul Affairs* **13**(3&4):133–165 (1996).

9. A. J. Bailer, Testing for the equality of area under the curves when using destructive measurement techniques. *J Pharmacokinet Biopharm* **16**:303–309 (1988).

10. C. Yeh, Estimation and significant tests of area under the curve derived from incomplete blood sampling. *ASA Proc* (*Biopharm Sec*) 74–81 (1990).

11. J. R. Nedelman, E. Gibiansky, and D. T. W. Lau, Applying Bailer's method for AUC confidence intervals to sparse sampling. *Pharm Res* **12**:124–128 (1995).

12. S. M. Pai, S. H. Fettner, G. Hajian, M. N. Cayen, and V. K. Batra, Characterization of AUCs from sparsely sampled populations in toxicology studies. *Pharm Res* **13**:1283–1290 (1996).

13. H.-M. Chu and E. I. Ette, A random sampling approach for robust estimation of tissue to plasma ratio from extremely sparse data. *AAPS J* **7**(1):E249–E258 (2005).

14. H. Mager and G. Goller, Resampling methods in sparse sampling situations in preclinical pharmacokinetic studies. *J Pharm Sci* **87**:372–378 (1008).

15. E. I. Ette and L. C. Onyiah, Estimating inestimable standard errors in population pharmacokinetic studies: the bootstrap with winsorization. *Eur J Drug Metab Pharmacokinet* **27**:213–224 (2002).

APPENDIX 42.1   CODE FOR NAIVE DATA AVERAGING APPROACH

```
####################################
# calculate AUC ratio by average data at each time point
# to construct the AUC curve by average conc. by time point
# source data set is named "data"
####################################


## calculate AUC ratio by average data by time point
## to construct the AUC curve by average conc. by time point
t_c(0,0.5,1,2,4,6,8)
a_apply(data.tissue,2,mean)
a_as.vector(a[-1])
b_apply(data.plasma,2,mean)
b_as.vector(b[-1])
```

```
# Trapezoidal.hmc is a subroutine (see 42.6.6)
c1_as.vector(Trapezoidal.hmc(time=t,DV=a))
c2_as.vector(Trapezoidal.hmc(time=t,DV=b))

## calculate ratio by time point
t_c(0.5,1,2,4,6,8)
a_rep(NA,length(t)+1)
tmp_data.frame(time=a,min=a,Q1=a,median=a,mean=a,Q3=a,max=a,NA=a,s
d=a,n=a)
for(i in 1:length(t))
{
    tmp[i,1]_as.character(t[i])
    print(t[i])
    print(summary(data$ratio[data$Time==t[i]]))
    b_summary(data$ratio[data$Time==t[i]])
    tmp[i,2:(length(b)+1)]_as.vector(b)
    tmp[i,9]_stdev(data$ratio[data$Time==t[i]])
    tmp[i,10]_length(data$ratio[data$Time==t[i]])

}

tmp[length(t)+1,1]_"overall"
b_summary(data$ratio)
tmp[length(t)+1,2:(length(b)+1)]_as.vector(b)
tmp[length(t)+1,9]_stdev(data$ratio)
tmp[length(t)+1,10]_length(data$ratio)
```

APPENDIX 42.2   CODE FOR RANDOM SAMPLING APPROACH

```
###############################################################
######
#### to impute AUC from scarified animals, each rat has one plasma and one
#### tissue data point

## to construct the plasma and tissue data sets and associated
column names (time points)
## // start
tmp_data.frame(rep=seq(1,irep),AUC.p=rep(NA,irep),AUC.
l=rep(NA,irep))
t_sort(unique(data$Time))
size.n_length(t)
data.plasma_data.frame(iter=c(1,2,3))
data.tissue_data.frame(iter=c(1,2,3))

for(i in 1:size.n)
{
  aa_data$plasma[data$Time==t[i]]
```

```
   data.plasma_cbind( data.plasma,aa)
   aa_data$tissue[data$Time==t[i]]
   data.tissue_cbind( data.tissue,aa)
}
dimnames(data.plasma)[[2]]_c("iter",paste("h",t,sep=""))
dimnames(data.tissue)[[2]]_c("iter",paste("h",t,sep=""))


## // end
## ------------------------------------------------------------

### to generate "irep" (M copies of) replication of Psuodoprofile and
### associated AUC for each subject in the data set


## // start


irep_1000

tmp.tissue_matrix(rep(NA,irep*size.n),ncol=size.n)
tmp.plasma_matrix(rep(NA,irep*size.n),ncol=size.n)

#data.plasma_data.plasma[,-c(1,2)]
#data.tissue_data.tissue[,-c(1,2)]

## irow*icol (18) is the loop for each subject
## icol is the loop for each time point within each subject
## j is the M-copy loop of calculating AUC for each subject
## the outcome stored in ans and renamed as ans.rep.**
ID.n_0
for(irow in 1:nrow(data.tissue))
{
   for(icol in 3:ncol(data.tissue))
   {
      # for each subject
      ID.n_ID.n+1

      tmp_data.frame(ID=rep(ID.n,irep),rep=seq(1,irep),AUC.
   p=rep(NA,irep),AUC.l=rep(NA,irep))
      tmp.tissue_matrix(rep(NA,irep*(size.n-1)),ncol=(size.n-1))
      tmp.plasma_matrix(rep(NA,irep*(size.n-1)),ncol=(size.n-1))
      for(i in 1:(size.n-1))
      {
        a_sample(x=c(1:3),size=irep,T)
        #cat("\n(",irow,icol,") -- with hour index",i,"\n")
        #print(a)
        #print(data.tissue[a,i+2])
        tmp.tissue [,i]_data.tissue[a,i+2]
        #print(tmp.tissue[,i] )
```

```
          tmp.plasma[,i]_data.plasma[a,i+2]
          #print(tmp.plasma[,i] )

   }

   for (j in 1:irep)
   {

       tmp$AUC.l[j]_Trapezoidal.hmc(time=t,DV=c(0,tmp.tissue[j,]))
       tmp$AUC.p[j]_Trapezoidal.hmc(time=t,DV=c(0,tmp.plasma[j,]))
       #cat("\n iter", irep,"AUC of tissue = ",tmp$AUC.l[j],"\n")
   }

   if ( icol*irow==3) {ans_tmp}
     else {ans_rbind(ans,tmp)}
 }
}

ans$ratio_ans$AUC.l/ans$AUC.p

summary(ans$ratio)
summary(ans$AUC.p)

boxplot(ans$ratio,ylab="ratio")
title("Distribution of Tissue to Plasma Ratio",cex=.9)

#ans.rep5_ans
#ans.rep10_ans
#ans.rep100_ans
#ans.rep500_ans

## // end
## -------------------------------------------------------------

boxplot(ans.rep5$ratio,ans.rep10$ratio,ans.rep100$ratio,ans.
rep500$ratio,
        names=c("5","10","100","500"),
  xlab="Replications for Each Subject",ylab="Tissue to Plasma Ratio")

#########################################################
#### part 2
#### to generate M copies of N=18 virual studies

data_ans.rep10
irep_50
n_18
#ans_matrix(rep(NA,n*irep),nrow=n) # row is the animal index and
  column is the replicates
ans_matrix(rep(NA,6*irep),nrow=irep) # columns are summary stats
  and row is the replicates
```

```
for (i in 1:irep)
{
  a_sample(x=seq(1:nrow(data)),size=n,T)
  #ans[,i]_data$ratio[a]
  ans[i,]_as.vector(summary(data$ratio[a]))
}

dimnames(ans)[[2]]_c("Min","Q1","Q2","Mean","Q3","Max")

#ans.rep5.summary_ans
#ans.rep10.summary_ans
#ans.rep100.summary_ans
#ans.rep500.summary_ans

par(mfrow=c(2,2))
boxplot(ans.rep5.summary[,3],ans.rep10.summary[,3],ans.rep100.
  summary[,3],ans.rep500.summary[,3],
     names=c("5","10","100","500"),
  xlab="replications",ylab="Tissue to Plasma Ratio",ylim=c(10,25))
title("Distribution of Median \n over 50 rep. of Virtual
Study",cex=.8)

boxplot(ans.rep5.summary[,4],ans.rep10.summary[,4],ans.rep100.
  summary[,4],ans.rep500.summary[,4],
     names=c("5","10","100","500"),
  xlab="replications",ylab="Tissue to Plasma Ratio",ylim=c(10,25))
title("Distribution of Mean \n over 50 rep. of Virtual
Study",cex=.8)

boxplot(ans.rep5.summary[,2],ans.rep10.summary[,2],ans.rep100.
  summary[,2],ans.rep500.summary[,2],
     names=c("5","10","100","500"),
  xlab="replications",ylab="Tissue to Plasma Ratio",ylim=c(10,25))
title("Distribution of Q1 \n over 50 rep. of Virtual
Study",cex=.8)

boxplot(ans.rep5.summary[,5],ans.rep10.summary[,5],ans.rep100.
  summary[,5],ans.rep500.summary[,5],
     names=c("5","10","100","500"),
  xlab="replications",ylab="Tissue to Plasma Ratio",ylim=c(10,25))
title("Distribution of Q3 \n over 50 rep. of Virtual
Study",cex=.8)

#################################
#### part 3
#### par.old_par()
frame()
par(oma=c(0,0,2,0),mar=c(5,5,4,3)+0.1)
```

```
par(mfrow=c(1,1))
par(fig=c(x1=0,x2=0.55,y1=0.45,y2=1))
boxplot(ans.rep5.summary[,3],ans.rep10.summary[,3],ans.rep100.
  summary[,3],ans.rep500.summary[,3],
      ans.rep1000.summary[,3],names=c("","","","",""),xlab="",ylab=
"",ylim=c(10,25))
mtext(side=3,"Median",line=1)


par(fig=c(x1=0.45,x2=1,y1=0.45,y2=1),yaxs="d")
boxplot(ans.rep5.summary[,4],ans.rep10.summary[,4],ans.rep100.
  summary[,4],ans.rep500.summary[,4],
      ans.rep1000.summary[,4],names=c("","","","",""),xlab="",ylab=
"",ylim=c(10,25),axes=F)
  box()
#title("Mean",cex=.7)
mtext(side=3,"Mean",line=1)


par(fig=c(x1=0,x2=0.55,y1=0,y2=0.55),xaxs="d",yaxs="d")
boxplot(ans.rep5.summary[,2],ans.rep10.summary[,2],ans.rep100.
  summary[,2],ans.rep500.summary[,2],
      ans.rep1000.summary[,2],names=c("5","10","100","500","1000"),
xlab="",ylab="",axes=T)
  box()
mtext(side=3,"1st Quartile",line=1)


par(fig=c(x1=0.45,x2=1,y1=0,y2=0.55),xaxs="d",yaxs="d")
boxplot(ans.rep5.summary[,5],ans.rep10.summary[,5],ans.rep100.
  summary[,5],ans.rep500.summary[,5],
      ans.rep500.summary[,5],names=c("5","10","100","500","1000"),x
      lab="",ylab="",axes=F)
box()
mtext(side=3,"3rd Quartile",line=1)


mtext("Distribution of Summary Stat. Over 50 Replications of Virtual
Study",outer=T)
mtext(side=1,"# of Replications for Each Subject",outer=T,line=-2)
```

APPENDIX 42.3   CODE FOR PSEUDOPROFILE-BASED BOOTSTRAP

```
############################################################
### to generate "irep" (M copies of) replication of Psuodoprofile and
### associated AUC for each subject in the data set
###
### n is the number of animals in the study
### irep == (B1 time loop in step 1 and 2 of the paper)
############################################################
```

```
b_c(10,50,seq(100,1000,100))

t_sort(unique(data$Time))
size.n_length(t)

for (k in 1:length(b))
{

  irep_b[k]
  n_18

  tmp.tissue_matrix(rep(NA,irep*size.n),ncol=size.n)
  tmp.plasma_matrix(rep(NA,irep*size.n),ncol=size.n)

  ## irow*icol (18) is the loop for each subject
  ## icol is the loop for each time point within irow-th subject
  ## j is the M-copy loop of calculating AUC for each subject
  ## the outcome stored in ans and rename as ans.rep.**

  # to sample by time point
  myStart_proc.time()

  for(i in 1:size.n)
  {
    a_sample(x=c(1:3),size=irep,T)
    tmp.tissue [,i]_data.tissue[a,i+1]
    tmp.plasma[,i]_data.plasma[a,i+1]
  }

  ## to calcualte AUC
  ans_data.frame(rep=rep(b[k],irep),AUC.l=rep(NA,irep),AUC.
p=rep(NA,irep))

  for ( i in 1:irep)
  {
    a_as.vector(unlist(tmp.tissue[i,]))
    ans$AUC.l[i]_Trapezoidal.hmc(time=t,DV=a)
    a1_as.vector(unlist(tmp.plasma[i,]))
    ans$AUC.p[i]_Trapezoidal.hmc(time=t,DV=a1)
  }
}

# to calculate the paired ratio
ans.PpbB.total$ratio.p_ans.PpbB.total$AUC.l/ans.PpbB.total$AUC.p

# to calculate the unpaired ratios
ans.PpbB.total$ratio.u_rep(NA,nrow(ans.PpbB.total))
for (i in 1:length(b))
#for (i in 1:1)
```

```
{
  b1_ans.PpbB.total$AUC.p[ans.PpbB.total$rep==b[i]]
  b2_ans.PpbB.total$AUC.l[ans.PpbB.total$rep==b[i]]
  b3_sample(x=c(1:length(b1)),size=length(b1),F)
  #print(b1)
  #print(b2)
  #print(b3)
  ans.PpbB.total$ratio.u[ans.PpbB.total$rep==b[i]]_b2/b1[b3]


}
```

APPENDIX 42.4   CODE FOR CONVERGENCE

```
#### compare the convergence among different M copies of paired ratios


##########################################################
############## random sampling approach

set.seed(565)
y_c(5,10,50,seq(100,1000,100)) # y is the # of replication for
each subject
# columns are summary stats and row is the replicates
ans1_matrix(rep(NA,3*length(y)),nrow=length(y))
ans2_matrix(rep(NA,3*length(y)),nrow=length(y))
ans3_matrix(rep(NA,3*length(y)),nrow=length(y))


for(k in 1:length(y))


{
  if(k==1){data_ans.rep5}
  if(k==2){data_ans.rep10}
  if(k>2 ){ data_ans.rep1000[ans.rep1000$rep<=y[k],]}
  irep_100
  n_18
  # columns are summary stats and row is the replicates
  ans_matrix(rep(NA,3*irep),nrow=irep)
  for (i in 1:irep)
  {
      a_sample(x=seq(1:nrow(data)),size=n,T)
      ans[i,1]_quantile(data$ratio[a],0.25)
      ans[i,2]_mean(data$ratio[a])
      ans[i,3]_quantile(data$ratio[a],0.75)

  }
  cat("\n",y[k],"\n")
```

```
    #print(ans)
    # calculate the 95% CI for the Q1
    ans1[k,1]_quantile(ans[,1],0.05)
    ans1[k,2]_mean(ans[,1])
    ans1[k,3]_quantile(ans[,1],0.95)

    # calculate the 95% CI for the mean
    ans2[k,1]_quantile(ans[,2],0.05)
    ans2[k,2]_mean(ans[,2])
    ans2[k,3]_quantile(ans[,2],0.95)

    # calculate the 95% CI for the Q3
    ans3[k,1]_quantile(ans[,3],0.05)
    ans3[k,2]_mean(ans[,3])
    ans3[k,3]_quantile(ans[,3],0.95)

}

## to save output for later computation purpose .rs = .random sam-
pling approach
    ans1.rs_ans1
    ans2.rs_ans2
    ans3.rs_ans3

#### to plot 95% CI for Q1,mean, and Q3 across different replica-
tions per subject
frame()
par(oma=c(0,0,2,0),mar=c(5,5,4,3)+0.1)
par(mfrow=c(1,3))
y_c(5,10,50,seq(100,1000,100)) # y is the # of replication for
each subject
x_seq(1:length(y))

par(fig=c(x1=0,x2=0.4,y1=0,y2=1))
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans1.rs[,2],lty=1)
lines(x=x,y=ans1.rs[,1],lty=2)
lines(x=x,y=ans1.rs[,3],lty=2)
axis(1, at=x,labels=as.character(y))
axis(2)
box()
mtext(side=3,"1st Quartile",line=1,cex=1.2)

par(fig=c(x1=0.3,x2=0.7,y1=0,y2=1))
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans2.rs[,2],lty=1)
lines(x=x,y=ans2.rs[,1],lty=2)
```

```
lines(x=x,y=ans2.rs[,3],lty=2)
axis(1,  at=x,labels=as.character(y))
#axis(2)
box()
mtext(side=3,"Mean",line=1,cex=1.2)

par(fig=c(x1=0.6,x2=1,y1=0,y2=1))
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans3.rs[,2],lty=1)
lines(x=x,y=ans3.rs[,1],lty=2)
lines(x=x,y=ans3.rs[,3],lty=2)
axis(1,  at=x,labels=as.character(y))
#axis(2)
box()
mtext(side=3,"3rd Quartile",line=1,cex=1.2)

mtext(side=1,"# of Replications for Each Subject",outer=T,line=-2,
cex=1.2)
mtext(side=2,"Tissue to Plasma AUC Ratio",outer=T,line=-2,cex=1.2)

############################################################
############### PpbB approach

set.seed(555)
y_c(10,50,seq(100,1000,100)) # y is the # of replication for each
subject
# columns are summary stats and row is the replicates
ans1_matrix(rep(NA,3*length(y)),nrow=length(y))
ans2_matrix(rep(NA,3*length(y)),nrow=length(y))
ans3_matrix(rep(NA,3*length(y)),nrow=length(y))

for(k in 1:length(y))

{

     data_ans.PpbB.rep1000[ans.PpbB.rep1000$rep<=y[k],]
     irep_100
     n_18
     ans_matrix(rep(NA,3*irep),nrow=irep)
#print("check pt 1")
#print(nrow(data))
     for (i in 1:irep)
     {
          a_sample(x=seq(1:nrow(data)),size=n,T)
#cat("\n",i,"rep",a,"\n")
#cat("\n",data$ratio.p[a],"\n")
     ans[i,1]_quantile(data$ratio.p[a],0.25)
```

```
#cat("\n","ans[i,1]",ans[i,1],"\n")
    ans[i,2]_mean(data$ratio.p[a])
    ans[i,3]_quantile(data$ratio.p[a],0.75)

    }
    cat("\n",y[k],"\n")
    #print(ans)
    # calculate the 95% CI for the Q1
    ans1[k,1]_quantile(ans[,1],0.05)
    ans1[k,2]_mean(ans[,1])
    ans1[k,3]_quantile(ans[,1],0.95)

    # calculate the 95% CI for the mean
    ans2[k,1]_quantile(ans[,2],0.05)
    ans2[k,2]_mean(ans[,2])
    ans2[k,3]_quantile(ans[,2],0.95)

    # calculate the 95% CI for the Q3
    ans3[k,1]_quantile(ans[,3],0.05)
    ans3[k,2]_mean(ans[,3])
    ans3[k,3]_quantile(ans[,3],0.95)

}

## to save output for later computation purpose .PpbB
    ans1.PpbB_ans1
    ans2.PpbB_ans2
    ans3.PpbB_ans3

#### to plot 95% CI for Q1,mean, and Q3 across different replica-
tions for each subject
frame()
par(oma=c(0,0,2,0),mar=c(5,5,4,3)+0.1)
par(mfrow=c(1,3))
y_c(10,50,seq(100,1000,100)) # y is the # of replication for each
subject
x_seq(1:length(y))

par(fig=c(x1=0,x2=0.4,y1=0,y2=1))
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans1.PpbB[,2],lty=1)
lines(x=x,y=ans1.PpbB[,1],lty=2)
lines(x=x,y=ans1.PpbB[,3],lty=2)
axis(1, at=x,labels=as.character(y))
axis(2)
box()
mtext(side=3,"1st Quartile",line=1,cex=1.2)

par(fig=c(x1=0.3,x2=0.7,y1=0,y2=1))
```

```
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans2.PpbB[,2],lty=1)
lines(x=x,y=ans2.PpbB[,1],lty=2)
lines(x=x,y=ans2.PpbB[,3],lty=2)
axis(1, at=x,labels=as.character(y))
#axis(2)
box()
mtext(side=3,"Mean",line=1,cex=1.2)

par(fig=c(x1=0.6,x2=1,y1=0,y2=1))
plot(x=0,y=0,type="n",ylim=c(10,25),xlim=range(x),xlab="",
  ylab="",axes=F)
lines(x=x,y=ans3.PpbB[,2],lty=1)
lines(x=x,y=ans3.PpbB[,1],lty=2)
lines(x=x,y=ans3.PpbB[,3],lty=2)
axis(1, at=x,labels=as.character(y))
#axis(2)
box()
mtext(side=3,"3rd Quartile",line=1,cex=1.2)

mtext(side=1,"Number of Bootstraps",outer=T,line=-2, cex=1.2)
mtext(side=2,"Tissue to Plasma AUC Ratio",outer=T,line=-2,cex=1.2)
```

APPENDIX 42.5   CODE FOR OUTLIER EFFECT

```
###################
#var.vec_c(0.1,0.2,0.3,0.4)  # variation percentage vector
var.vec_c(0.1)
missing.vec_c(1)             # vector of outlier occurrence
myRep_50                     # replication of one scenario, e.g. 1
missing with 10% outlier

a.n_length(var.vec)*length(missing.vec)*myRep
a_rep(NA,a.n)
myAns_data.frame(n.outlier=a,cv.outlier=a,pooling=a,PpbB=a,RS=a)

B1_100
irep_100

a.n_0
for(i in 1:length(var.vec))
{
  for (j in 1:length(missing.vec))
  {
     cat("\n","CV =",var.vec[i]*100,"% n.of.outliers =",j,"\n")
     for (k in 1:myRep)
```

```
    {
        a.n_a.n+1
        myVar_var.vec[i]  # variation percentage
                myCount_missing.vec[j]  # number of outliers
                myAns$n.outlier[a.n]_myCount
                myAns$cv.outlier[a.n]_myVar


        cat("\n",k," pool average")

    #-- to construct data set with outlier/s and calculated the
pooled average

    myData0_outlier.sub(liverOrNot=1,data=study2385,myVar=myVar,
myCount=myCount)
    myAns$pooling[a.n]_myData0$r.pooling

    cat(" PpbB")

    #-- use pseudoprofile approach

    myData_PpbB.sub1(B1=B1,data=myData0$data,tissue=myData0$
tissue,plasma=myData0$plasma,n=18)

    myData2_PpbB.sub2(B1=B1,data=myData,irep=irep,n=18)
    myAns$PpbB[a.n]_mean(myData2$Q2)

    cat(" Random Sampling")

                #-- use random sample approach

    myData3_rs.sub1(tissue=myData0$tissue,plasma=myData0$plasma,
irep=B1)
    myData4_rs.sub2(data=myData3,irep=irep,n=18)

                myAns$RS[a.n]_mean(myData4$Q2)


        }
    }
}

###################################
# graph with one outlier across 4 levels of increase
frame()
par(oma=c(0,0,2,0),mar=c(5,5,4,3)+0.1)
par(mfrow=c(2,2))

par(fig=c(x1=0,x2=0.55,y1=0.45,y2=1))
```

```
tmp_myAns.all[myAns.all$n.outlier==1 & myAns.all$cv.outlier==0.1,]
boxplot(tmp$pooling,tmp$PpbB,tmp$RS,names=c("","",""),ylim=c(14,20
))
title("(a) 10%",cex=0.6)


par(fig=c(x1=.45,x2=1,y1=0.45,y2=1))
tmp_myAns.all[myAns.all$n.outlier==1 & myAns.all$cv.outlier==0.2,]
boxplot(tmp$pooling,tmp$PpbB,tmp$RS,names=c("","",""),ylim=c(14,20
),axes=F)
title("(b) 20%",cex=0.6)
box()


par(fig=c(x1=0,x2=0.55,y1=0,y2=.55))
tmp_myAns.all[myAns.all$n.outlier==1 & myAns.all$cv.outlier==0.3,]
boxplot(tmp$pooling,tmp$PpbB,tmp$RS,names=c("Naive","PpbB","RS"),y
lim=c(14,20))
title("(c) 30%",cex=0.6)


par(fig=c(x1=.45,x2=1,y1=0,y2=.55))


tmp_myAns.all[myAns.all$n.outlier==1 & myAns.all$cv.outlier==0.4,]
boxplot(tmp$pooling,tmp$PpbB,tmp$RS,names=c("Naive","PpbB","RS"),
ylim=c(14,20),axes=F)
title("(d) 40%",cex=0.6)
box()


mtext(side=2,"Tissue to Plasma AUC Ratio",out=T,cex=.9,line=-1.5)


mtext(side=1,"Method",out=T,cex=.9,line=-1.5)
```

APPENDIX 42.6   CODE FOR OTHER SUBROUTINES

```
##############################################################
#########
#### file name : outlier Effect subroutine

#----------------------------------------------------------------
outlier.sub_function(tissueOrNot=1,data,myVar,myCount=)
{
  #tissueOrNot_1   # outlier/s location: 1 is tissue, 2 is plasma
  #myVar_.2        # variation percentage
  #myCount_2       # number of outliers
  #data_study2385[,1:3]
  n_nrow(data)
  data$row_seq(1:n) # row position

  x_data$row[data$Time>0 & data$Time<4]  # x is targeted row vector
```

```
  # ramdomly select # of "myCount" from 18 "x" subjects, who will
be replaced
  #with outlier/s
     a_sample(x=x,size=myCount,F)
     data$row[a]_1000  # to indicate outlier replacement positions
     b_1

     c_1+b*myVar
     if(tissueOrNot==1)
         {data$tissue[a]_data$tissue[a]*c}
     else
         {data$plasma[a]_data$plasma[a]*c}

## to construct the plasma and tissue data sets and associated
column names (time points)

     t_sort(unique(data$Time))
     size.n_length(t)
     data.plasma_data.frame(iter=c(1,2,3))
     data.tissue_data.frame(iter=c(1,2,3))

     for(i in 1:size.n)
     {
         aa_data$plasma[data$Time==t[i]]
         data.plasma_cbind( data.plasma,aa)
         aa_data$tissue[data$Time==t[i]]
         data.tissue_cbind( data.tissue,aa)
     }
     dimnames(data.plasma)[[2]]_c("iter",paste("h",t,sep=""))
     dimnames(data.tissue)[[2]]_c("iter",paste("h",t,sep=""))

  #### to calculate the AUC ratio by mean pooling
  # the ratio is saved as r.pooling

  a_apply(data.plasma,2,mean)
  a_as.vector(unlist(a))
  a_a[-1]
  b_as.vector(Trapezoidal.hmc(time=t,DV=a) )

  a_apply(data.tissue,2,mean)
  a_as.vector(unlist(a))
  a_a[-1]
  b1_as.vector(Trapezoidal.hmc(time=t,DV=a) )

  r.pooling_b1/b
  ans_list(data=data,r.pooling=r.pooling,tissue=data.
tissue,plasma=data.plasma)
  invisible(ans)
}
```

```
#------------------------------------------------------------------

### to generate "B1" (M copies of) replication of Psuodoprofile and
### associated AUC for each subject in the data set

PpbB.sub1_function(B1,data,tissue,plasma,n)
{

     ## // start
     # n is the number of animals in the study
     ## irep == (B1 time loop in step 1 and 2 of the paper)

     t_sort(unique(data$Time))
     size.n_length(t)

     for (k in 1:length(B1))
     {

          irep_B1[k]
          #n_18

          junk.tissue_matrix(rep(NA,irep*size.n),ncol=size.n)
          junk.plasma_matrix(rep(NA,irep*size.n),ncol=size.n)

          ## irow*icol (18) is the loop for each subject
          ## icol is the loop for each time point within irow-th
             subject
          ## j is the M-copy loop of calculating AUC for each
             subject
          ## the outcome stored in ans and rename as ans.rep.**

          # to sample by time point

          for(i in 1:size.n)
          {

               a_sample(x=c(1:3),size=irep,T)
               junk.tissue [,i]_tissue[a,i+1]
               junk.plasma[,i]_plasma[a,i+1]
          }

  ## to calcualte AUC

  ans_data.frame(rep=rep(B1[k],irep),AUC.l=rep(NA,irep),AUC.
p=rep(NA,irep))

  for ( i in 1:irep)
  {
```

```
      a_as.vector(unlist(junk.tissue[i,]))
      ans$AUC.l[i]_Trapezoidal.hmc(time=t,DV=a)
      a1_as.vector(unlist(junk.plasma[i,]))
      ans$AUC.p[i]_Trapezoidal.hmc(time=t,DV=a1)

  }
  if(k==1){ans.PpbB.total_ans}
  else {ans.PpbB.total_rbind(ans.PpbB.total,ans)}

  }

  # to calculate the paired ratio
  ans.PpbB.total$ratio.p_ans.PpbB.total$AUC.l/ans.PpbB.total$AUC.p

  # to calculate the unpaired ratios
  ans.PpbB.total$ratio.u_rep(NA,nrow(ans.PpbB.total))
  for (i in 1:length(B1))
  #for (i in 1:1)

  {
    b1_ans.PpbB.total$AUC.p[ans.PpbB.total$rep==B1[i]]
    b2_ans.PpbB.total$AUC.l[ans.PpbB.total$rep==B1[i]]
    b3_sample(x=c(1:length(b1)),size=length(b1),F)
    ans.PpbB.total$ratio.u[ans.PpbB.total$rep==B1[i]]_b2/b1[b3]

  }
  invisible(ans.PpbB.total)
}

#-------------------------------------------------------------------
PpbB.sub2_function(B1=300,data=ans.PpbB.total,irep=100,n=18)
{
  ans_matrix(rep(NA,6*irep),nrow=irep) # columns are summary stats
and row is the replicates
  for (i in 1:irep)
      {
          a_sample(x=seq(1:nrow(data)),size=n,T)
          a_sample(x=seq(1:nrow(data)),size=n,T)
          #ans[,i]_data$ratio.p[a]
          ans[i,]_as.vector(summary(data$ratio.p[a]))

      }
  ans_as.data.frame(ans)
  dimnames(ans)[[2]]_c("Min","Q1","Q2","Mean","Q3","Max")
  invisible(ans)

}
```

```
#------------------------------------------------------------------

##############################################
### random sampling approach

rs.sub1_function(tissue,plasma,irep=300)
{

    junk.tissue_matrix(rep(NA,irep*size.n),ncol=size.n)
    junk.plasma_matrix(rep(NA,irep*size.n),ncol=size.n)
    ## irow*icol (18) is the loop for each subject
    ## icol is the loop for each time point within each subject
    ## j is the M-copy loop of calculating AUC for each subject
    ## the outcome stored in ans and rename as ans.rep.**
    ID.n_0
    for(irow in 1:nrow(tissue))
    {
        for(icol in 3:ncol(tissue))
        {
        # for each subject
        ID.n_ID.n+1

    junk_data.frame(ID=rep(ID.n,irep),rep=seq(1,irep),AUC.
p=rep(NA,irep),AUC.l=rep(NA,irep))
        junk.tissue_matrix(rep(NA,irep*(size.n-1)),ncol=(size.n-1))
        junk.plasma_matrix(rep(NA,irep*(size.n-1)),ncol=(size.n-1))
        for(i in 1:(size.n-1))
        {
            a_sample(x=c(1:3),size=irep,T)
            junk.tissue [,i]_tissue[a,i+2]
            junk.plasma[,i]_plasma[a,i+2]
        }

        for (j in 1:irep)
        {

    junk$AUC.l[j]_Trapezoidal.hmc(time=t,DV=c(0,junk.tissue[j,]))

    junk$AUC.p[j]_Trapezoidal.hmc(time=t,DV=c(0,junk.plasma[j,]))
    #cat("\n iter", irep,"AUC of tissue = ",junk$AUC.l[j],"\n")
        }

        if ( icol*irow==3) {ans_junk}

        else {ans_rbind(ans,junk)}
    }
  }
```

```
  ans$ratio_ans$AUC.l/ans$AUC.p
  invisible(ans)
}


#-----------------------------------------------------------------


##########################################################
#### random sampling approach
## to generate M copies of N=18 virual studies

rs.sub2_function(data,irep=100,n=18)
{
    ans_matrix(rep(NA,6*irep),nrow=irep) # columns are summary stats
and row is the replicates

    for (i in 1:irep)
    {
        a_sample(x=seq(1:nrow(data)),size=n,T)
        #ans[,i]_data$ratio[a]
        ans[i,]_as.vector(summary(data$ratio[a]))
    }
    ans_as.data.frame(ans)
    dimnames(ans)[[2]]_c("Min","Q1","Q2","Mean","Q3","Max")
    invisible(ans)
}


#-----------------------------------------------------------------


Trapezoidal.hmc_function(time,DV)
{
    if(length(time)!=length(DV))
      {stop("The counts of DV and Time are different!")}
    else
    {
        n_length(time)
        x_time[2:n]-time[1:(n-1)]
        y_(DV[2:n]+DV[1:(n-1)])/2
        a_t(x)%*%y
        a
        invisible(a)
    }
}
```

Physiologically Based Pharmacokinetic Modeling: Inhalation, Ingestion, and Dermal Absorption

SASTRY S. ISUKAPALLI, AMIT ROY, and PANOS G. GEORGOPOULOS

43.1   INTRODUCTION

In recent years, physiologically based pharmacokinetic (PBPK) modeling has been increasingly adopted in the pharmaceutical industry due to the feasibility of developing these models without data from in vivo experiments. A significant part of the PBPK modeling literature to date deals with toxicological risk assessment, and the models have often been developed using data from high-dose experiments on animals. In the pharmaceutical industry, empirical "compartmental" pharmacokinetic models (hereafter referred to simply as PK models) have been used far more widely than their PBPK counterparts.

Both the PBPK and PK models can be classified as transport–transformation models, as they describe the transport and metabolism (transformation) of chemicals within physiological systems; these are key processes affecting concentrations of chemicals in tissues, and hence the responses of physiological systems to exposure.

A key difference between PK and PBPK models is that the PBPK models can be used to predict concentration–time profiles in several tissues of interest, whereas the predictions of a PK model are constrained to tissues accessible to measurement. This is because the PK models are "phenomenological," whereas PBPK models are "mechanistic." More specifically, though both PK and PBPK models describe the body in terms of compartments, in the case of PK models, the compartments need not represent any physiological entities, whereas in PBPK models, the compartments represent organs or tissues or groups of organs and tissues (1). The structure of a PBPK model is derived from basic anatomical and physiological structure of the organism studied, whereas the structure of a PK model is typically based on the available drug disposition data. This can be considered as a major distinctive feature and advantage of PBPK models.

Another advantage of PBPK models over PK models is that PBPK models are more amenable to different types of rational, mechanistically based extrapolation, including cross-species, cross-tissue, cross-chemical, and high-dose to low-dose extrapolations. Since PK models are empirical, they should strictly be used for interpolation only within the range of the experimental data. The cross-species extrapolation of PBPK models is based on the rationale that the whole body structure is similar in different mammals, such as mice, rats, dogs, and humans. In fact, a majority of PBPK models in toxicology and risk assessment have been developed by utilizing animal experiments and have been scaled for risk assessment studies in humans. Because PK models are empirical, the model parameters are often estimated using in vivo response data, and the number of parameters in these models is limited by the identifiability of these parameters with respect to available response data. In contrast, PBPK models are mechanistically based and have physically meaningful parameters, most of which can be obtained from independent experiments or from the literature. A majority of remaining PBPK model parameters can be estimated from in vitro data without the need for data from expensive in vivo studies. Furthermore, PBPK models are also amenable to iterative refinement based on new data, often collected from independent experiments (1, 2).

The main limitation of PBPK models is that the number of model parameters is usually much higher than the corresponding PK models. Therefore, parameter estimation in case of PBPK models is a more complex task. Though PBPK models are amenable to iterative refinement via utilization of parameter estimates from multiple sources, significantly more computational and mathematical expertise is required in developing a PBPK model than the corresponding PK model (1). Moreover, because of the large number of model parameters in PBPK models, sufficient effort must be invested in performing comprehensive sensitivity and uncertainty analyses of these models.

Overall, PBPK models can provide insight into the several aspects associated with the kinetics of a drug within the human body, collectively termed as ADMET, for absorption, distribution, metabolism, elimination, and toxicity. An application of the PBPK models at the early stage of drug development can be useful to rapidly screen candidate drugs based on their PK properties via in silico approaches (3, 4). Due to the rapid increases in the computational power, and the parallel advances in the PBPK area, the role of PBPK models in pharmacometrics is likely to substantially increase.

This chapter serves three purposes: (a) to provide a brief overview of PBPK modeling, (b) to present a tutorial on the issues and steps involved in the development of a PBPK model, and (c) to present an application and discuss relevant issues associated with model refinement, evaluation, parameter estimation, and sensitivity/uncertainty analysis. First, some basic background information is provided, and references to important resources are presented. Then the process of developing a PBPK model is discussed, and a step-by-step description of a PBPK modeling example is provided, along with a brief discussion on relevant complementary issues such as model parameter estimation and sensitivity/uncertainty analysis. The example is presented in a manner that a novice PBPK modeler can follow the model structure, mathematical equations, and the code. Relevant cross-references between the equations, parameter tables, and the actual code is presented. Though the example is implemented in Matlab (5), it does not require substantial Matlab

experience to run the code and make minor changes to the code. The example is aimed to facilitate even beginner programmers to easily follow the code.

43.2   OVERVIEW OF PBPK MODELING

Though the conceptual formulation of PBPK models can be attributed to Teorell (6, 7), computational PBPK models were first implemented by Bischoff and Brown (8) and have been extensively applied to describe the pharmacokinetics of toxic chemicals. An extensive overview of the PBPK modeling studies for a wide range of chemicals is presented by Reddy et al. (2), with a compilation of past PBPK studies including volatile organics, aromatics, pesticides, dioxins, metals, and chemical mixtures. The classification is based primarily on the class of chemicals, so a researcher can easily obtain a list of different models and approaches used for a given chemical. One reason that PBPK models are the tool of choice for estimating the concentration–time profiles of toxic chemicals in the body is that it is generally unethical to conduct experiments on humans with these chemicals. The PBPK modeling paradigm offers the advantage of a rational means of predicting the pharmacokinetics of a chemical in humans, based on the pharmacokinetics in animals. It also offers the possibility of refining the characterization of human pharmacokinetics based on in vitro data obtained from human-derived experimental systems. Furthermore, PBPK models also provide a rational means of estimating the exposure of one or more target organs to the toxic chemical.

A useful overview of PBPK modeling from the perspective of the pharmaceutical industry has been published by Nestorov (1), focusing on the applicability of PBPK modeling for medicinal compounds. Another overview of the recent progress in PBPK modeling pertinent to drug development and regulatory science is provided by Rowland et al. (9). PBPK modeling issues relevant to drugs are also discussed by Reddy (10), with a chapter devoted to an overview of PBPK modeling for antineoplastic agents.

43.3   STEPS IN FORMULATING A PBPK MODEL

The main steps in the formulation of a PBPK model include specifying the mathematical structure, specifying model parameter values, and computational (software) implementation of the model. In cases where in vivo data are available, some of the model parameters may be estimated from the in vivo data using statistical parameter estimation techniques. Several resources are available to aid scientists interested in using existing PBPK models or implementing new PBPK models. These resources include knowledge bases of PBPK modeling literature, and databases of relevant physiological parameters and biochemical properties. Additionally, computational toolkits for rapid development and implementation of PBPK models and supplementary toolkits for PBPK parameter estimation and uncertainty analysis are also available. A knowledge base of relevant resources for PBPK modeling is presented at the Physiomics section of the Environmental Bioinformatics Knowledge Base (ebKB; `www.environmentalbioinformatics.org`).

43.3.1   Specifying Structure of the PBPK Model

The mathematical formulation of the PBPK model is dependent on several factors: routes of intake of a chemical or sites of drug administration, target tissues of interest, physiological components to be explicitly modeled (kinetically important tissues and organs and the linkages between them), transport processes of the chemical (flow, diffusion, disposition, clearance, etc.), and metabolic processes involved.

- The main routes of intake of a chemical are ingestion (dietary or nondietary), inhalation, dermal absorption, and parenteral (intravenous, intramuscular, intrathecal, intraperitoneal, and subcutaneous). The structure of a PBPK model is dependent on the intake routes, as the corresponding organs or tissues usually need to be explicitly modeled in order to describe the uptake of the chemical. It is therefore advisable to identify routes of uptake prior to developing the PBPK model.
- The choice of organs and tissues to be explicitly modeled in the PBPK model depends on several factors: sites of dose administration and exposure scenarios; sites of metabolic activity within the body; target tissues or potential sites of action; and physicochemical (thermodynamic) attributes of the chemical, such as lipophilicity. Organs that are usually explicitly described in a PBPK model include liver (primary site of biotransformation), lung (for volatile chemicals that are absorbed by inhalation and eliminated by exhalation), skin (for exposure scenarios that include dermal absorption), gastrointestinal tract (for absorption of ingested chemicals), and kidney (for renal excretion). Both arterial and venous bloodstreams are usually modeled separately (either explicitly as compartments or implicitly) and are linked to other compartments in a manner representative of body blood circulation. The remaining tissues and organs are usually lumped into three compartments according to their kinetic characteristics: "rapidly perfused," "slowly perfused," and "fat." The kinetic characteristics of a compound and thus the choice of lumping scheme are dependent on the chemical properties of the compound (lipophilic/hydrophilic) and the rate of blood flow to the tissue, relative to its volume. Ideally, a sensitivity analysis should be performed by considering competing model structures, and an appropriate model should be selected based on the sensitivity analysis. An organ should be explicitly modeled and not lumped with kinetically similar tissue provided that a sensitivity analysis indicated that doing so results in a meaningful change in the target organ exposure. Figure 43.1 shows the PBPK model structure used in this example application.
- The main processes governing the pharmacokinetics of a chemical are absorption, distribution, metabolism, and excretion. In PBPK models, distribution of a chemical is characterized by blood flow rates to each organ and tissue, and partitioning of the chemical between tissue and blood. These processes are commonly modeled using two alternative types of assumptions: flow-limited and diffusion-limited transport. The flow-limited assumption implies that equilibration between free and bound fractions in blood and tissue is rapid, and that concentrations of the chemical in the venous blood exiting a tissue and in the tissue are at equilibrium. The tissue is assumed to be a homogeneous

FIGURE 43.1 Schematic of the structure of the PBPK model structure used in the example. (Adapted from Roy et al. (46).)

well-mixed space with respect to concentration of the chemical. On the other hand, a diffusion-limited assumption implies that there are significant barriers to transport of the chemical between blood and tissue. In the diffusion-limited case, permeability rates are used to describe transport of chemical between blood and the tissue. It is also possible to have diffusion-limited transport within a tissue (e.g., between intracellular and interstitial spaces). In such a case, the tissue is divided into multiple compartments, each of which is well mixed, and diffusion-limited permeability rates are used for transport across these compartments. Usually, in the diffusion-limited case, the tissue is modeled as two subcompartments with diffusion-limited transport between them (11). Complex tissue models taking into account multiple components

within a tissue have also been proposed (12). The main limitation of tissue models with multiple compartments is the large number of parameters that are required. In some cases diffusional barriers may exist to transport across the entire tissue, not just between subspaces within a tissue. In this case the well-mixed tissue assumption is not appropriate, and the concentration gradient of the chemical within the tissue is described by a partial differential equation. The concentration gradient is characterized by a single parameter, the diffusion constant. An example is shown in the PBPK application described in Section 43.5.

· The mathematical structure of a PBPK model typically involves a set of ordinary differential equations. Partial differential equations can be solved by discretizing the compartment containing a concentration gradient into multiple identical compartments, each of which can be represented by an ordinary differential equation. An example is the discretization of the stratum corneum subcompartment of the skin, shown in Section 43.5.

43.3.2 Specifying PBPK Model Parameters

Once the structure of the PBPK model is formulated, the next step is specifying the model parameters. These can be classified into a chemical-independent set of parameters (such as physiological characteristics, tissue volumes, and blood flow rates) and a chemical-specific set (such as blood/tissue partition coefficients, and metabolic biotransformation parameters). Values for the chemical-independent parameters are usually obtained from the scientific literature and databases of physiological parameters. Specification of chemical-specific parameter values is generally more challenging. Values for one or more chemical-specific parameters may also be available in the literature and databases of biochemical and metabolic data. Values for parameters that are not expected to have substantial interspecies differences (e.g., tissue/blood partition coefficients) can be imputed based on parameter values in animals. Parameter values can also be estimated by conducting in vitro experiments with human tissue. Partitioning of a chemical between tissues can be obtained by vial equilibration or equilibrium dialysis studies, and metabolic parameters can be estimated from in vitro metabolic systems such as microsomal and isolated hepatocyte sytems. Parameters not available from the aforementioned sources can be estimated directly from in vivo data, as discussed in Section 43.4.5.

43.3.2.1 Physiological Parameters
Typical values for the physiological parameters such as tissue volumes, blood flow rates, inhalation rates, and body surface area (for dermal absorption) can be obtained from the literature (13–15). For a specific target group of individuals, allometric scaling can be used based on the available physiological data for the individuals to be studied (such as age, gender, and body weight) (16). For population PK models that characterize the interindividual variability in pharmacokinetics among subjects in a population, distributions of the parameters in the population are needed. It should be noted that it is generally not appropriate to assume that the parameters distributions are independent. A database of physiological parameter distributions that accounts for correlations is provided as part of the Physiological Parameters for

PBPK Modeling (P3M) program (17). The joint probability distribution of relevant physiological parameters can be constructed empirically from the P3M database, conditioned on age, sex, and ethnicity.

43.3.2.2 Partition Coefficients

Biochemical parameters such as tissue/blood partition coefficients can be obtained from in vitro experiments with vial equilibration (18) or equilibrium dialysis techniques (19). A less expensive process involves exploiting the similarities in physical characteristics of similar tissues in animals and humans and using in vitro animal data (20). Other, even more cost-effective techniques include extrapolation from experimentally determined octanol/water partition coefficients (21), or in silico parameter estimation via techniques based on structure–property relationships (22, 23).

For common toxic chemicals, the Agency for Toxic Substances and Disease Registry (ATSDR) (`www.atsdr.cdc.gov/toxpro2.html`) maintains "toxicological profiles" with detailed information about toxicokinetic parameters for the chemical, with available reports for over 250 chemicals. Several commercial ADMET programs include databases of partition coefficients and modules for estimating them for a large set of chemicals. Examples include the KnowItAll informatics system from BioRad (`www.knowitall.com`) and chemical property predictor software from ChemSilico (`www.chemsilico.com`).

43.3.2.3 Biotransformation Parameters

Similar to the techniques used for calculation of chemical disposition parameters, in vivo biotransformation kinetic parameters of a substrate can be estimated from in vitro systems such as microsomes, freshly isolated hepatocytes, liver slices, and isolated perfused livers (24). Intrinsic clearance or Michaelis–Menten parameters for the whole liver can also be obtained by "scaling" in vitro parameters based on the cytochrome P450 enzyme content (25–27). These parameters can also be estimated from in vitro data obtained from recombinant human CYP systems, and also through allometric scaling of clearance estimates from animal PBPK models.

43.4 STEPS IN MODEL IMPLEMENTATION, EVALUATION, AND REFINEMENT OF PBPK MODELS

43.4.1 Resources for Implementation of the PBPK Model

The computational implementation of a PBPK model can be accomplished using a variety of software tools. Available tools for PBPK modeling range from general purpose computational modeling systems to interfaces specifically designed for PBPK modeling. Since the PBPK models are typically described by sets of ordinary differential equations, it is possible to implement a PBPK model in any modern programming language or modeling system with relative ease by using the available numerical libraries. Some of the commonly used tools for PK and PBPK modeling are shown in Table 43.1. Comparative evaluation studies using some of these tools for PBPK simulation have been presented in the literature (28, 29). A comprehensive list of resources for the implementation of the PBPK models is available in the

TABLE 43.1   List of Software Systems Useful for the Development and Analysis of PBPK Models

Software Systems	Reference	Notes
Environments for Rapid Generation of PBPK Models		
ERDEM (exposure related dose estimating model)	`www.epa.gov/heasdweb/`	Allows graphical PBPK model design and provides parameters for a set of toxic chemicals. Requires the free advanced continuous simulation language (ACSL) viewer from `www.acslsim.com`
PKQuest	`www.pkquest.com`	Contains a set of PBPK modeling routines in Maple and requires Maple (commercial; `www.maplesoft.com`)
WinSAAM	`www.winsaam.com`	Based on modeling and linking compartments; graphical design of PBPK
acslXtreme PK/PD Toolkit (commercial)	`www.aegisxcellon.com/ Pharmacokinetic _Toolkit.html`	Built in PK, PBPK, and PD blocks, and features such as Monte Carlo and sensitivity analysis
ModKine (commercial)	`www.biosoft.com/ w/modkine.htm`	General kinetics modeling program with PK/PD support
PK-Sim (commercial)	`www.pk-sim.com`	PBPK simulation via proprietary modules for the calculation of model parameters from easily obtainable chemical properties
Simcyp (commercial)	`www.simcyp.com`	Graphical, interactive environment for PK modeling, including enzyme kinetics, and drug–drug interactions
Gastro-Plus	`http://www.simulations -plus.com/products/ gastro_plus/`	Graphical program simulating absorption and pharmacokinetics of orally dosed drugs
General Purpose Scientific Modeling Tools		
Matlab (commercial)	`www.mathworks.com`	Powerful scripting language with general purpose toolboxes, and graphical model design using SIMULINK; also provides user-contributed PBPK toolboxes
GNU Octave	`www.octave.org`	A free alternative to Matlab
acslXtreme (commercial)	`www.aegisxcellon.com`	Scientific modeling system with graphical model design, and a PK/PD toolkit. Supports the ACSL language

TABLE 43.1   *Continued*

Software Systems	Reference	Notes
WinNolnlin (commercial)	`www.pharsight.com/ products/winnonlin`	General purpose simulation system with a set of PK, PD, and PK/PD link models
STELLA (commercial)	`www.hps-inc.com`	Provides an icon-based interface for model building and simulation, and also for sensitivity analysis
Berkeley Madonna (commercial)	`www.berkeleymadonna.com`	General purpose differential equation solver
ADAPT II	`bmsr.usc.edu/Software/`	Simulation system for models described by first-order ODEs (includes some PK/PD examples)
General Data Analysis and Parameter Estimation Tools		
WinNonMix (commercial)	`www.pharsight.com/ products/winnonmix`	Interactive system for nonlinear mixed-effects modeling
NONMEM (commercial)	`www.globomax.com`	General purpose toolkit for parameter estimation for nonlinear mixed effects models (NONMEM).
MCSim	`toxi.ineris.fr/en/`	Provides interfaces for designing ODE models and utilize Bayesian inference through the Markov chain Monte Carlo (MCMC) approach
BUGS (Bayesian inference using Gibbs sampling)	`mathstat.helsinki.fi/ openbugs/`	Bayesian analysis of complex statistical models using MCMC

Physiomics section of the Environmental Bioinformatics KnowledgeBase (ebKB; `www.environmentalbioinformatics.org`).

43.4.2   Testing, Evaluation, and Refinement of the PBPK Model

Once a PBPK model is developed and implemented, it should be tested for mass balance consistency, as well as through simulated test cases that can highlight potential errors. These test cases often include "software boundary conditions," such as zero dose and high initial tissue concentrations. Some parameters in the PBPK model may have to be estimated through available in vivo data via standard techniques such as nonlinear regression or maximum likelihood estimation (30). Furthermore, in vivo data can be used to update existing (or prior) PBPK model parameter estimates in a Bayesian framework, and thus help in the refinement of the PBPK model. The Markov chain Monte Carlo (MCMC) (31–34) is one of the

most widely used Bayesian techniques for parameter estimation. Once the model parameters are estimated from available data, techniques such as the posterior predictive checking (35, 36) can be utilized to assess the model performance.

43.4.3   Sensitivity and Uncertainty Analysis

Comprehensive sensitivity analyses of PBPK models, focusing on the sensitivity of model outcomes to both the model structure and model parameters, can provide guidelines toward selecting appropriate model structures and can direct data gathering needs. Conventionally, the sensitivity of model outputs to model inputs is described by the sensitivity coefficients, typically the partial derivatives of the model output with respect to each input (37), either through multiple simulations, or using the inbuilt features of the software packages. For example, the Advanced Continuous Simulation Language (ACSL) system (www.acslsim.com) provides functions for automatically calculating the sensitivity coefficients of model outputs. Some of the advanced techniques for performing sensitivity analysis are based on computer processing of the PBPK model source code. One example is the automatic differentiation technique, which has been applied for sensitivity and uncertainty analysis of PBPK models (38).

Characterizing the overall uncertainties associated with the PBPK model estimates is also an important component of the PBPK model evaluation and application. This includes characterizing the uncertainties in model outputs resulting from the uncertainty in the PBPK model parameters. Traditionally, Monte Carlo has been employed for performing uncertainty analysis of PBPK models (39, 40). Some of the recent techniques that have been applied for the uncertainty analysis of PBPK models include the stochastic response surface method (SRSM) (38, 41) and the high-dimensional model reduction (HDMR) technique (42).

43.4.4   Extending the Model for Population Studies (Variability Analysis)

PBPK models for individuals can be extended to estimate the variability of response in a population by addressing the interindividual variability in model parameters. Ideally, the variability in PBPK model parameters is specified by a joint probability distribution that accounts for correlations between each pair of model parameters. However, the substantial amount of data that would be needed to develop the joint distribution is often not available. A first approximation of correlations between physiological parameters in the model can be obtained by ensuring that joint distributions of body weight, age, and gender are maintained. The model parameters that are functions of these individual covariates will then reflect these correlations. A convenient source of individual covariates is provided by the P3M program (17). A practical method for estimating the joint distribution of all model parameters from in vivo data is provided by empirical Bayesian methods using Markov chain Monte Carlo (MCMC) simulation, discussed in the following section.

43.4.5   Model Parameter Estimation

Estimation or refinement of model parameters, sometimes described as the "inverse problem," is complex because such problems usually do not have a unique

solution, and the solution methods are often concerned with finding the "best possible solution" with respect to some objective criteria. One of the commonly used techniques for parameter estimation is maximum likelihood estimation (MLE). Given the model parameter set and data, the MLE technique is based on maximizing the likelihood (conditional probability, given a set of parameter values) that differences between data and model results are due to random error. The MLE approach has been used to estimate metabolic parameter values (V_{max}, and K_m) from in vivo data (43), but the approach cannot be used to simultaneously estimate all the parameters in the model, as all the parameters are not simultaneously identifiable (44). Bayesian techniques, such as the Markov chain Monte Carlo (MCMC) (31, 32) simulation, provide probabilistic estimates of an arbitrarily large number of parameters based on prior parameter information in the form of probability distributions and experimental data and are being applied increasingly in PBPK modeling (33, 34, 45).

43.5 APPLICATION EXAMPLE: A PBPK MODEL FOR CHLOROFORM

An example of a PBPK model that illustrates some of the concepts described in the preceding sections is presented below. This example is based on a PBPK model of chloroform in humans from Roy et al. (46) as presented in Figure 43.1. Chloroform is a water disinfection (chlorination) by-product present in municipal tap water and in swimming pools, and relatively short-term dermal and ingestion exposures to chloroform occur during showering, bathing, and swimming. The general approach described here is directly applicable to the modeling of several gaseous and volatile organics. Furthermore, the structure of the PBPK model shown here can be adapted in a relatively straightforward manner to nonvolatile chemicals as well.

The following section presents a step-by-step tutorial example of the implementation of a PBPK model for chloroform. The model has been coded in Matlab Version 7.0, and effort is made to design it in the form of a reference implementation that is easily extensible for PBPK models for other chemicals. A summary of the Matlab code is provided in Table 43.4, and code listings are presented in the appendix.

43.5.1 Model Structure

The chloroform PBPK model describes intake from inhalation, dermal absorption from either water or air, and oral ingestion from drinking water. The intravenous injection route is also added in this model so that it can be used for other chemicals. The corresponding media concentrations of chloroform, in the air, drinking water, and shower water, are dependent on the exposure scenario and hence are to be provided as model inputs.

Liver, kidney, skin, gut, and fat are explicitly modeled as individual compartments, with the remaining tissues and organs lumped into either "slowly perfused" or "rapidly perfused" tissue. The lung is also modeled as a separate compartment but is used primarily to describe the exchange of chloroform between the air and blood, with the assumption that the venous blood and air exiting the lung compartment are in equilibrium with respect to concentration of chemical. It is assumed that a negligible amount of the chemical accumulates in the lung, which can be justified

on the grounds that the mass of the lung is a small fraction of total body mass, and therefore any accumulation is likely to be negligible.

The main assumption used with respect to the biological transport of the chemical is that of "flow-limited" transport. Inhalation is modeled by assuming that the inhaled air mixes instantaneously and reaches equilibrium with the venous blood in the lung. Dermal absorption is modeled using three approaches: one compartment representing the skin (one-skin), two compartments representing the skin (two-skin), and a distributed compartment representing the skin (distributed-skin), as shown in Figure 43.2. Oral ingestion via drinking water is modeled by assuming that the all the chloroform in drinking water is rapidly absorbed in GI lumen.

The tissue volume fractions are assumed to be fixed, but the actual total tissue volume is based on sex-dependent allometric scaling factors and body weight. The cardiac flow rates to different tissues are also based on allometric scaling factors. However, customizations based on other assumptions (e.g., sex-dependent tissue volume fractions) or on physiological parameter databases can easily be performed by modification of the code presented in Code Listing 3 (e.g., Appendix 43.3, Lines 45, 46, 99, etc.).

FIGURE 43.2   The three different models for the skin compartment used in the PBPK model. (Adapted from Roy et al. (46).)

Chloroform is assumed to be metabolized in either the liver or the kidney, and the rates of biotransformation in these tissue are described by Michaelis–Menten kinetics. Metabolism is implemented in a general manner, with biotransformation possible in any compartment. Biotransformation is limited to the liver and kidney by setting the V_{max} coefficients for compartments other than the liver and kidney to zero (see Lines 100 and 101 in Code Listing 3, Appendix 43.3).

43.5.2 Specifying Model Inputs and Outputs

Model inputs consist of exposure scenario and physiological characteristics of the exposed individual. An exposure scenario is specified by chloroform concentrations in media (air and water) and the duration of contact with water. Media concentrations can change with time, depending on the location and activity of the individual. Physiological characteristics of an individual, such as body weight, surface area, age, and sex, are used as inputs into the model. The main model outputs of interest are exhaled air concentration–time profile, area under concentration–time profile in liver (AUCL), and cumulative amount metabolized in liver (CML). Exhaled air concentrations serve as a surrogate for arterial blood concentrations, as these concentrations are assumed to be in equilibrium. AUCL and CML are calculated as measures of exposure, because the liver is the main target organ of chloroform toxicity, and the toxicity is believed to be mediated by a reactive metabolite (47, 48).

43.5.3 Mathematical Formulation

The mathematical formulation of the PBPK model is derived by applying mass balance rules across multiple compartments. The general form of mass balance equations is the same for the fat, slowly perfused, rapidly perfused, and kidney compartments, whereas the mass balance equations for the liver, gut, and lung compartments are unique.

Table 43.2 shows the general model parameters (excluding dermal) used in this modeling example, and Table 43.3 shows the dermal model parameters. For all the equations, cross-references are provided to the relevant lines of the Matlab code

TABLE 43.2 PBPK Model Parameters Used in the Example Application, Excluding the Dermal Model Parameters

Parameter/ Variable	Description	Value	Code Variable/Line Number (Cxx:Lyy)
BW	Body weight (kg)	70	indiv.BW; (C1:L7)
D	Average density of body	1.0	c.Density; (C3:L37)
K_m	Michaelis-Menten constant (µg/L) in liver	448	c.MMK.KM_C; (C3:L89)
$Q_{cardiac}$	Total cardiac output (L/min) derived from BW and sex based allometric factor	6.40	c.Q_cardiac; (C3:L99)
$Q_{alveolar}$	Inhalation rate (L/min) (scenario based)	5.80	events.rate_inhale; (C1:L20)

TABLE 43.2   *Continued*

Parameter/ Variable	Description	Value	Code Variable/Line Number (Cxx:Lyy)
V_{max}	Maximum rate of metabolism (µg/min) derivedfrom allometric scaling factor and BW	5120.62	c.VMAX (C3:101); Scaled from c.MMK.VMAX_C (C3:L88)

Partition Coefficients (PC)

$P_{blood:air}$	Blood:air PC	7.43	c.PC.blood_air; (C3:L78)
$P_{fat:blood}$	Fat:blood PC	37.7	c.PC.tissue(ID.Fat); (C3:L79)
$P_{sp:blood}$	Slowly perfused tissue: blood PC	1.62	c.PC.tissue(ID.SP); (C3:L1.80)
$P_{gut:blood}$	Gut:blood PC	2.3	c.PC.tissue(ID.Gut); (C3:L1.82)
$P_{rp:blood}$	Rapidly perfused tissue: blood PC	2.3	c.PC.tissue(ID.RP); (C3:L81)
$P_{liv:blood}$	Liver:blood PC	2.3	c.PC.tissue(ID.Liv); (C3:L83)
$P_{kid:blood}$	Kidney:blood PC	1.5	c.PC.tissue(ID.Kid); (C3:L84)

Fractional Blood Flow Rates (QF) to Compartments (L/min)

QF_{fat}	QF to fat	0.05	c.QCR(ID.Fat); (C3:L49)
QF_{liv}	QF to liver	0.07	c.QCR(ID.Liv); (C3:L52)
QF_{gut}	QF to gut	0.18	c.QCR(ID.Gut); (C3:L53)
QF_{rp}	QF to rapidly perfused tissue	0.26	c.QCR(ID.RP); (C3:L51)
QF_{kid}	QF to kidney	0.25	c.QCR(ID.Kid);(C3:L54)
QF_{skin}	QF to skin	0.04	c.QCR(ID.Skin);(C3:L55)
QF_{sp}	QF to slowly perfused tissue; autoadjusted so that sum of all fractions equals 1	0.15	c.QCR(ID.SP); (C3:L50) Auto adjust: (C3:L56)

Fractional Tissue Volumes (VF)
(scaled to obtain actual volumes based on body weight and density)

VF_{fat}	VF of fat	0.231	c.VR(ID.Fat); (C3:L60)
VF_{venous}	VF of venous blood (arbitrarilyselected)	0.01	c.venous_blood_volume_fraction; (C3:L58)
VF_{liv}	VF of liver	0.0314	c.VR(ID.Liv); (C3:L63)
VF_{rp}	VF of rapidly perfused tissue	0.0327	c.VR(ID.RP); (C3:L62)
VF_{kid}	VF of kidney	0.0044	c.VR(ID.Kid); (C3:L65)
VF_{gut}	VF of gut	0.017	c.VR(ID.Gut); (C3:L64)
VF_{skin}	VF of skin	0.1	c.VR(ID.Skin); (C3:L66)
VF_{sp}	VF of slowly perfused tissue; auto adjusted so that sum of all fractions equals 1	0.5735	c.VR(ID.SP); (C3:L61) Auto adjust: (C3:L69)

TABLE 43.3   *Continued*

Parameter/ Variable	Description	Value	Code Variable/Line (Cxx:Lyy)
A	Body surface area (cm^2)	1800	`indiv.BSA; (C1:L8)`
$K_{media:skin}$ (w)	Overall skin permeability in water (cm/min)	1.84×10^{-4}	`c.K.skin_water;` `(C4:L54)`
$K_{media:skin}$ (a)	Overall skin permeability in air (cm/min) (computed via $K_{media:skin}$ (w))	9.085×10^{-4}	`c.K.skin_air;` `(C4:L58)`
$P_{skin:blood}$	Skin:blood PC	1.6	`c.PC.tissue(ID.Skin);` `(C4:L53)`
$P_{skin:air}$	skin:air PC (from skin:blood and blood:air PCs)	11.888	`c.PCMC.skin_air;` `(C4:L56)`
$P_{skin:water}$	Skin:water PC (from skin:air and air:water PCs)	1.9377	`c.PCMC.skin_water;` `(C4:L57)`
$P_{sc:water}$	Sc:water PC (from octanol:water PC)	24.4069	`c.PCMC.sc_water;` `(C4:L83)`
$P_{sc:air}$	sc:air PC (from sc: water and air: water PCs)	149.7354	`c.PCMC.sc_air;` `(C4:L84)`
$P_{vs:sc}$	vs:sc PC (from vs: water, sc:water, skin:water, etc.)	0.077	`c.PCMC.vs_sc;` `(C4:L85-L89)`
$P_{vs:blood}$	vs:blood PC (from skin:water and sc:water PCs, etc.)	1.5522	`c.PCMC.vs_blood;` `(C4:L88)`
$K_{media:sc}$ (w)	sc permeability in water (cm/min)	1.4809×10^{-4}	`c.K.sc_water;` `(C4:L93)`
$K_{media:sc}$ (a)	sc permeability in air (cm/min) (from water $K_{media:sc}$)	9.085×10^{-4}	`c.K.sc_air;` `(C4:L94)`
$K_{sc:vs}$	sc:vs permeability (from sc:water PC and Cleek and Bunge's B parameter calculation [46])	0.0098	`c.K.vs_sc;` `(C4:L95)`
D_{sc}	effective sc diffusivity (cm^2/min) (from sc:water and L_sc)	6.0676×10^{-9}	`c.DM.D_sc;` `(C4:L107)`
L_{sc}	sc thickness (cm)	0.001	`c.D.L_sc; (C4:L60)`

Miscellaneous Parameters Used in the Model

	Octanol:water coefficient	90	`c.PC.chem.octanol_` `water; (C4:L80)`

TABLE 43.4   Overview of the Matlab Code for the Chloroform PBPK Model Example

File Name; Code Listing and Lines	Description
PBPK_chloroform_case_study.m Code Listing 1 (53 Lines)	Presents the main case study showing how the model configuration and exposure events can be set up for different model structures (e.g., one-skin vs. distributed-parameter skin models) and for different exposure scenarios (e.g., inhalation and dermal, inhalation only, and dermal only)
PBPK_voc_driver.m Code Listing 2 (107 Lines)	The "main" simulation function that solves the ordinary differential equations that describe the PBPK model. The function returns outputs at different time steps and the configuration information in a structure. It also provides the mass balances at the corresponding time steps
PBPK_chloroform_config.m Code Listing 3 (138 Lines)	Describes the basic structure of the PBPK model used, including the compartments and state variables (based on model options), allometric scaling factors, partition coefficients, and metabolic constants
PBPK_add_demal_model.m Code Listing 4 (110 Lines)	Updates the PBPK model configuration by adding the dermal component, based on the model options provided. This module provides an example of making further customizations to the PBPK model configuration. This model is invoked by the main configuration function.
PBPK_voc_deriv.m Code Listing 5 (66 Lines)	Describes the derivatives of the state variables in the model as a function of time. This is invoked indirectly through Matlab's ordinary differential equation solver. This model, along with the configuration module, describes the core aspects of the PBPK model
PBPK_add_dermal_derivative.m Code Listing 6 (72 Lines)	Updates the derivatives with the dermal component. This module also provides an example of making further customizations to the PBPK model structure and equations. This function is invoked by the main derivative function.

included, in order to facilitate better understanding of the formulation and implementation of the PBPK model; the cross-references are in the form "Cxx:Lyy", where "xx" indicates the Code Listing number, and "yy" indicates the Line Number within that Code Listing.

The mass balance equation for the fat, slowly perfused, rapidly perfused, and kidney compartments is shown in Eq. (43.1).

$$V_j \frac{dc_j}{dt} = Q_j \left(c_{\text{arterial}} - c_{v,j} \right) - R_j \quad (\text{C5}:\text{L40})$$

$$c_{v,j} = \frac{c_j}{P_{j:\text{blood}}} \quad\quad\quad\quad (\text{C5}:\text{L26})$$

$$R_j = \frac{V_{\max,j} \cdot c_j}{K_{m,j} + c_j} \quad\quad\quad (\text{C5}:\text{L39})$$

$$Q_j = QF_j \cdot Q_{\text{cardiac}} \quad\quad (\text{C5}:\text{L106})$$

$$V_j = VF_j \cdot \text{BW} \cdot \text{D} \quad\quad (\text{C5}:\text{L97})$$

(43.1)

where c_j is the concentration of the chemical in the jth compartment (with c_{arterial} denoting the concentration in the arterial blood), Q_j is the blood flow to the jth compartment, V_j is the volume of the jth compartment, $c_{v,j}$ is the concentration in the blood exiting the jth compartment, and $V_{\max,j}$ and $K_{m,j}$ are the Michaelis–Menten constants for the jth compartment. The rest of the symbols used in Eq. (43.1) are described in Table 43.2, for different compartments.

For the gut compartment, the amount ingested is assumed to be absorbed into the GI lumen based on the bioavailability fraction, which is assumed to be equal to 1 (C3:L105). Thus, for a constant rate of ingestion, the equations for gut and lumen can be described by Eqs. (43.1) and (43.2).

$$V_{\text{gut}} \frac{dc_{\text{gut}}}{dt} = Q_{\text{gut}} \left(c_{\text{arterial}} - c_{v,\text{gut}} \right) - R_{\text{gut}} + T_{\text{gut:lumen}} \quad (\text{C5}:\text{L40, L54})$$

$$\frac{da_{\text{lumen}}}{dt} = S_{\text{ingest}} \bullet f_{\text{bioavail}} - T_{\text{gut:lumen}} \quad\quad (\text{C5}:\text{L52, L53})$$

$$T_{\text{gut:lumen}} = K_{\text{gut:lumen}} \bullet a_{\text{lumen}} \quad\quad (\text{C5}:\text{L51})$$

(43.2)

Here, f_{bioavail} is the fraction of chemical absorbed into the body and is assumed to be 1.0 (C3:L105), $K_{\text{gut:lumen}}$ is the transfer rate of the chemical from the lumen to gut (C3:L91), a_{lumen} is the amount of chemical in the lumen, and S_{ingest} is the rate of ingestion of the chemical.

Bolus injection events are handled by updating the corresponding state variables, such as the amount of chemical in venous blood for intravenous bolus injection, and amount of chemical in GI lumen for bolus ingestion, as shown in Code Listing 2 (Appendix 43.2), Lines 45–51.

The concentration of chloroform in the liver compartment is described by Eq. (43.1) and (43.3).

$$V_{\text{liver}} \frac{dc_{\text{liver}}}{dt} = Q_{\text{liver}} \left(c_{\text{arterial}} - c_{v,\text{liver}} \right) + Q_{\text{gut}} \left(c_{v,\text{gut}} - c_{v,\text{liver}} \right) + R_{\text{liver}}$$

(43.3)

$$(\text{C5}:\text{L40, L42, L43})$$

The arterial blood concentration is governed by Eq. (43.4), which is based on the equilibrium mass balance for the lung compartment.

$$c_{\text{arterial}} = \frac{Q_{\text{cardiac}}\, c_{\text{venous}} + Q_{\text{alveolar}}\, c_{\text{air(inhaled)}}}{Q_{\text{cardiac}} + Q_{\text{alveolar}}/P_{\text{blood:air}}} \qquad (\text{C3}:\text{L34, L35})$$

$$V_{\text{venous\_blood}} \frac{dc_{\text{venous}}}{dt} = Q_{\text{cardiac}}(c_{\text{venous\_mixed}} - c_{\text{venous}}) + S_{\text{infusion}} \quad (\text{C5}:\text{L45, L46})$$

$$c_{\text{venous\_mixed}} = \frac{1}{Q_{\text{cardiac}}}\left(Q_{\text{gut}} c_{v,\text{liver}} + \sum_{\substack{j=1 \\ j \neq \text{lung,gut}}}^{n} Q_j c_{v,j} \right) \qquad (\text{C3}:\text{L28, L30, L31})$$

$$(43.4)$$

where n is the total number of compartments, and the summation is performed for all compartments, excluding the lung and gut. $V_{\text{venous\_blood}}$ is the amount of blood in a hypothetical "venous blood compartment," where the infusion is assumed to take place, which is assumed to be 1% of the body volume (C3:L58). S_{infusion} is the rate of infusion of the chemical into the venous bloodstream, and c_{venous} is the concentration leaving the venous blood compartment after mixing and infusion, which is the same as the concentration of the chemical in the blood entering the lung compartment.

The mass balance equation for the concentration in the skin compartment is dependent on the choice of the skin model.

For the one-compartment skin model, the concentration in the skin compartment is given by Eq. (43.5).

$$V_{\text{skin}} \frac{dc_{\text{skin}}}{dt} = Q_{\text{skin}}(c_{\text{arterial}} - c_{v,\text{skin}}) + A \cdot K_{\text{media:skin}}\left(c_{\text{media}} - \frac{c_{\text{skin}}}{P_{\text{skin:media}}} \right)$$

$$c_{v,\text{skin}} = \frac{c_{\text{skin}}}{P_{\text{skin:blood}}} \quad (\text{C6}:\text{L31}-\text{L133}; \text{C5}:\text{L26})$$

$$(43.5)$$

For the two-compartment skin model, the concentration in the stratum corneum (c_{sc}) and in viable skin (c_{vs}) can be calculated using Eq. (43.6).

$$V_{\text{sc}} \frac{dc_{\text{sc}}}{dt} = A \cdot K_{\text{media:sc}}\left(c_{\text{media}} - \frac{c_{\text{sc}}}{P_{\text{sc:media}}} \right) - A \cdot K_{\text{sc:vs}}\left(c_{\text{sc}} - \frac{c_{\text{vs}}}{P_{\text{vs:sc}}} \right)$$

$$(\text{C6}:\text{L37}-\text{L41, L42})$$

$$V_{\text{vs}} = \frac{dc_{\text{vs}}}{dt} Q_{\text{skin}}\left(c_{\text{arterial}} - \frac{c_{\text{vs}}}{P_{\text{skin:blood}}} \right) + A \cdot K_{\text{sc:vs}}\left(c_{\text{sc}} - \frac{c_{\text{vs}}}{P_{\text{vs:sc}}} \right)$$

$$(\text{C6}:\text{L37}-\text{L41, L43})$$

$$(43.6)$$

For the distributed-parameter skin compartment model, the concentration c_{sc} is calculated by discretizing the stratum corneum compartment into a set of $N + 2$ equidistant nodes and using the central difference formula. This results in the representation of the one-dimensional Fickian diffusion equation to calculate mass flux at any depth within the stratum corneum (46):

$$J(x,t) \equiv -D_{\text{sc}} \frac{\partial c_{\text{sc}}(x,t)}{\partial x}$$

resulting in

$$V_{\text{sc}} \frac{dc_{\text{sc},i}}{dt} = \frac{D_{\text{sc}}}{[L_c/(N+1)]^2}(c_{\text{sc},i+1} - 2c_{\text{sc},i} + c_{\text{sc},i-1}); \quad i = 1,\ldots,N \quad (\text{C6}:\text{L61}) \qquad (43.7)$$

where $c_{sc,i}$ is the concentration in the ith node. The boundary conditions for this system are

$$c_{sc,0} = c_{media} \cdot P_{sc:media} \quad \text{and} \quad c_{sc,N+1} = c_{vs}/P_{vs:sc} \quad (C6:L52-L55) \tag{43.8}$$

The concentration in the viable skin compartment is given by

$$V_{vs}\frac{dc_{vs}}{dt} = Q_{skin}\left(c_{arterial} - \frac{c_{vs}}{P_{skin:blood}}\right) - A \cdot D_{sc}\left(\frac{c_{sc,N+1} - c_{sc,N}}{L_c/(N+1)}\right) \quad (C6:L67) \tag{43.9}$$

43.5.4 Computational Model Implementation

The model has been implemented in Matlab 7.0 and utilizes features for defining flexible structures (objects) in Matlab in order to improve both the model readability and extensibility. All the parameters that are needed to specify the model are included in a model configuration object, which is specified as part of the model configuration section.

Minor modifications to the structure of the model (such as changing the number of tissue compartments) can be made by modifying code in the configuration module, without having to change the code that specifies the differential equations. The configuration module specifies model parameters, based on the characteristics of the individual (age, sex, and body weight).

43.5.5 Model Application and Scenario Simulation

In this example, two chloroform exposure scenarios were simulated: (a) dermal-only and (b) simultaneous inhalation and dermal exposures. The inhalation exposures were assumed to occur from breathing air containing chloroform at $0.1\,\mu g/L$, whereas the dermal exposures were assumed to occur due to skin contact with water containing chloroform at $150\,\mu g/L$. In both scenarios, the exposure duration was assumed to be 30 minutes. The exposure event details are presented in Lines 13–22 of Code Listing 1 (Appendix 43.1). Depending on the type of exposures, the corresponding changes were made prior to solving the model (e.g., on Line 33 of Code Listing 1, the air concentrations were set to zero, in order to simulate the "dermal-only" exposure scenario). Two types of model formulations were studied here: the one-compartment skin model and the distributed-parameter skin model. The initial concentrations of chloroform in all tissues were assumed to be zero.

Figure 43.3 shows the exhaled breath concentration profiles estimated by the one-compartment and distributed-parameter skin models for a dermal-only exposure, whereas Figure 43.4 shows the estimates for a combined inhalation and dermal exposure. In both cases, the one-compartment skin model estimates higher peak values, and both models predict similar values at steady-state conditions. In the dermal-only exposure scenario, only the distributed-parameter skin model captures the lag between the dermal exposure and exhaled breath concentrations. For the combined inhalation and dermal exposure scenario, the distributed-parameter skin model shows a rise in the exhaled breath concentrations after the end of the exposure; this can be attributed to chloroform in the skin slowly diffusing into the

FIGURE 43.3   Simulated exhaled breath concentrations resulting from a 30 min dermal-only exposure to chloroform (150 μg/L water concentration), as estimated by two different model formulations for the skin compartment.

FIGURE 43.4   Simulated exhaled breath concentrations resulting from a 30 min inhalation and dermal exposure to chloroform (150 μg/L water concentration, and 0.1 μg/L air concentration), as estimated by two different model formulations for the skin compartment.

FIGURE 43.5 Estimates of the cumulative amount of chloroform metabolized in the liver (CML) resulting from a 30 min dermal-only exposure to chloroform (150 μg/L water concentration), as estimated by two different model formulations for the skin compartment.

FIGURE 43.6 Estimates of the area under concentration–time profile in liver (AUCL) resulting from a 30 min dermal-only exposure to chloroform (150 μg/L water concentration), as estimated by two different model formulations for the skin compartment.

FIGURE 43.7 Estimates of the cumulative amount of chloroform metabolized in the liver (CML) resulting from a 30 min inhalation and dermal exposure to chloroform (150 μg/L water concentration, and 0.1 μg/L air concentration), as estimated by two different model formulations for the skin compartment.

bloodstream. The one-compartment skin model does not show this rise because the entire skin compartment is assumed to be in instantaneous equilibrium with the blood.

Figures 43.5 and 43.6 show the cumulative amount metabolized in the liver (CML) and area under concentration–time profile in liver (AUCL) for a dermal-only exposure. In this case, the one-compartment skin model estimates higher CML and AUCL relative to the distributed-parameter skin model and does not account for the lag between the start of the dermal exposure and choloroform appearing in the bloodstream. However, after the end of the exposure, estimates from both models gradually converge. Similary, Figures 43.7 and 43.8 show the CML and the AUCL for a combined inhalation and dermal exposure. In this exposure scenario, AUCL and CML curves diverge following the end of the inhalation exposure, with the one-compartment skin model predicting higher values for both the CML and the AUCL. In both exposure scenarios, a "time lag" can be observed between the two models, and the one-compartment skin model estimates higher CML and AUCL relative to the distributed-parameter skin model, with the CML and AUCL curves gradually converging after the end of the exposure.

43.6   SUMMARY

An overview of PBPK modeling is presented here focusing on the step-by-step development of a PBPK model, followed by a reference implementation of a PBPK model for chloroform. The model formulation is general in nature and can be extended to nonvolatile chemicals. The example presented here aims to facilitate

FIGURE 43.8 Estimates of the area under concentration–time profile in liver (AUCL) resulting from a 30 min inhalation and dermal exposure to chloroform (150 μg/L water concentration, and 0.1 μg/L air concentration), as estimated by two different model formulations for the skin compartment.

an easy transition of a traditional PK modeler into PBPK modeling, addressing both from the mathematical and computational aspects. The mathematical formulation and the Matlab code in the example present a tutorial introduction to the computational implementation of a PBPK model, without requiring significant Matlab programming experience. A scientist who is a beginner in programming as well as in PK modeling should be able to easily follow the equations and the code provided in this chapter, and should be able to gain an understanding of multiple aspects of PBPK model development and implementation.

ACKNOWLEDGMENT

REFERENCES

1. I. Nestorov, Whole body pharmacokinetic models. *Clin Pharmacokinet* **42**:883–908 (2003).
2. M. B. Reddy, R. S. H. Yang, H. J. Clewell III, and M. E. Andersen, *Physiologically Based Pharmacokinetic Modeling: Science and Applications.* Wiley, Hoboken, NJ, 2005, pp. 392–393.

3. M. Bertrand, P. Jackson, and B. Walther, Rapid assessment of drug metabolism in the drug discovery process. *Eur J Pharm Sci* **11**:S61–S72 (2000).

4. A. Boobis, U. Gundert-Remy, P. Kremers, P. Macheras, and O. Pelkonen, In silico prediction of ADME and pharmacokinetics—report of an expert meeting organised by COST B15. *Eur J Pharm Sci* **17**:183–193 (2002).

5. S. R. Otto and J. P. Denier, *An Introduction to Programming and Numerical Methods in MATLAB.* Springer, New York, 2005.

6. T. Teorell, Kinetics of distribution of substances administered to the body. II: The intravascular mode of administration. *Arch Int Pharmacodyn* **57**:226–240 (1937).

7. T. Teorell, Kinetics of distribution of substances administered to the body. I: The extravascular modes of administration. *Arch Int Pharmacodyn* **57**:205–225 (1937).

8. K. B. Bischoff and R. G. Brown, Drug distribution in mammals. *Chem Eng Prog* **62**:32–45 (1966).

9. M. Rowland, L. Balant, and C. Peck, Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, May 29–30, 2002). *AAPS PharmSci* **6**:E6 (2004).

10. M. B. Reddy, *Physiologically Based Pharmacokinetic Modeling: Science and Applications.* Wiley, Hoboken, NJ, 2005.

11. S. Bjorkman, D. R. Wada, D. R. Stanski, and W. F. Ebling, Comparative physiological pharmacokinetics of fentanyl and alfentanil in rats and humans based on parametric single-tissue models. *J Pharmacokinet Biopharm* **22**:381–410 (1994).

12. C. Tanaka, R. Kawai, and M. Rowland, Physiologically based pharmacokinetics of cyclosporine A: reevaluation of dose–nonlinear kinetics in rats. *J Pharmacokinet Biopharm* **27**:597–623 (1999).

13. J. Fisher, M. Lumpkin, J. Boyd, D. Mahle, J. Bruckner, and H. A. El-Masri, PBPK modeling of the metabolic interactions of carbon tetrachloride and tetrachloroethylene in B6C3F1 mice. *Environ Toxicol Pharmacol* **16**:93–105 (2004).

14. J. W. Fisher, D. Mahle, and R. Abbas, A human physiologically based pharmacokinetic model for trichloroethylene and its metabolites, trichloroacetic acid and free trichloroethanol. *Toxicol Appl Pharmacol* **152**:339–359 (1998).

15. J. C. Lipscomb, J. W. Fisher, P. D. Confer, and J. Z. Byczkowski, In vitro to in vivo extrapolation for trichloroethylene metabolism in humans. *Toxicol Appl Pharmacol* **152**:376–387 (1998).

16. S. L. Lindstedt and P. J. Schaeffer, Use of allometry in predicting anatomical and physiological parameters of mammals. *Lab Anim* **36**:1–19 (2002).

17. P. S. Price, R. B. Conolly, C. F. Chaisson, E. A. Gross, J. S. Young, E. T. Mathis, and D. R. Tedder, Modeling interindividual variation in physiological factors used in PBPK models of humans. *Crit Rev Toxicol* **33**:469–503 (2003).

18. T. Kaneko, P. Y. Wang, and A. Sato, Partition-coefficients of some acetate esters and alcohols in water, blood, olive oil, and rat-tissues. *Occup Environ Med* **51**:68–72 (1994).

19. C. Ottiger and H. Wunderli-Allenspach, Partition behaviour of acids and bases in a phosphatidylcholine liposome–buffer equilibrium dialysis system. *Eur J Pharm Sci* **5**:223–231 (1997).

20. H. B. Zhang, A new approach for the tissue–blood partition coefficients of neutral and ionized compounds. *J Chem Inf Modeling* **45**:121–127 (2005).

21. M. P. Payne and L. C. Kenny, Comparison of models for the estimation of biological partition coefficients. *J Toxicol Environ Health Part A* **65**:897–931 (2002).

22. S. Selinski, K. Golka, H. M. Bolt, and W. Urfer, Estimation of toxicokinetic parameters in population models for inhalation studies with ethylene. *Environmetrics* **11**:479–495 (2000).

23. R. A. Corley, S. M. Gordon, and L. A. Wallace, Physiologically based pharmacokinetic modeling of the temperature-dependent dermal absorption of chloroform by humans following bath water exposures. *Toxicol Sci* **53**:13–23 (2000).

24. S. A. Wrighton, M. Vandenbranden, J. C. Stevens, L. A. Shipley, B. J. Ring, A. E. Rettie, and J. R. Cashman, In-vitro methods for assessing human hepatic drug-metabolism—their use in drug development. *Drug Metab Rev* **25**:453–484 (1993).

25. K. Venkatakrishnan, L. L. von Moltke, R. S. Obach, and D. J. Greenblatt, Drug metabolism and drug interactions: application and clinical value of in vitro models. *Curr Drug Metab* **4**:423–459 (2003).

26. B. A. Hoener, Predicting the hepatic-clearance of xenobiotics in humans from in-vitro data. *Biopharm Drug Dispos* **15**:295–304 (1994).

27. J. B. Houston and A. Galetin, Progress towards prediction of human pharmacokinetic parameters from in vitro technologies. *Drug Metab Rev* **35**:393–415 (2003).

28. C. E. Staatz and S. E. Tett, Comparison of two population pharmacokinetic programs, NONMEM and P-PHARM, for tacrolimus. *Eur J Clin Pharmacol* **58**:597–605 (2002).

29. M. R. Easterling, M. V. Evans, and E. M. Kenyon, Comparative analysis of software for physiologically based pharmacokinetic modeling: simulation, optimization, and sensitivity analysis. *Toxicol Methods* **10**:203–229 (2000).

30. E. I. Ette and P. J. Williams, Population pharmacokinetics II: estimation methods. *Ann Pharmacother* **38**:1907–1915 (2004).

31. D. J. Lunn, N. Best, A. Thomas, J. Wakefield, and D. Spiegelhalter, Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn* **29**:271–307 (2002).

32. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* Chapman & Hall, Boca Raton, FL, 1998.

33. F. Jonsson and G. Johanson, The Bayesian population approach to physiological toxicokinetic–toxicodynamic models—an example using the MCSim software. *Toxicol Lett* **138**:143–150 (2003).

34. F. Y. Bois, Statistical analysis of Fisher et al. PBPK model of trichloroethylene kinetics. *Environ Health Perspect* **108**:275–282 (2000).

35. Y. Yano, S. L. Beal, and L. B. Sheiner, Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn* **28**:171–192 (2001).

36. B. Green and S. B. Duffull, Prospective evaluation of a D-optimal designed population pharmacokinetic study. *J Pharmacokinet Pharmacodyn* **30**:145–161 (2003).

37. M. V. Evans, W. D. Crank, H. M. Yang, and J. E. Simmons, Applications of sensitivity analysis to a physiologically based pharmacokinetic model for carbon tetrachloride in rats. *Toxicol Appl Pharmacol* **128**:36–44 (1994).

38. S. S. Isukapalli, A. Roy, and P. G. Georgopoulos, Efficient sensitivity/uncertainty analysis using the combined stochastic response surface method and automated differentiation: application to environmental and biological systems. *Risk Anal* **20**:591–602 (2000).

39. W. J. Cronin, E. J. Oswald, M. L. Shelley, J. W. Fisher, and C. D. Flemming, A trichloroethylene risk assessment using a Monte-Carlo analysis of parameter uncertainty in conjunction with physiologically-based pharmacokinetic modeling. *Risk Analy* **15**:555–565 (1995).

40. H. J. Clewell, J. M. Gearhart, P. R. Gentry, T. R. Covington, C. B. VanLandingham, K. S. Crump, and A. M. Shipp, Evaluation of the uncertainty in an oral reference dose for methylmercury due to interindividual variability in pharmacokinetics. *Risk Anal* **19**:547–558 (1999).

41. S. S. Isukapalli, A. Roy, and P. G. Georgopoulos, Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems. *Risk Anal* **18**:351–363 (1998).

42. S. W. Wang, P. G. Georgopoulos, G. Li, and H. Rabitz, Characterising uncertainties in human exposure modelling through the random sampling–high dimensional model representation (RS-HDMR) methodology. *Int J Risk Assess Manage* **5**:387–406 (2005).

43. A. E. Smith and J. S. Evans, Uncertainty in fitted estimates of apparent in-vivo metabolic constants for chloroform. *Fundam Appl Toxicol* **25**:29–44 (1995).

44. F. Jonsson, F. Bois, and G. Johanson, Physiologically based pharmacokinetic modeling of inhalation exposure of humans to dichloromethane during moderate to heavy exercise. *Toxicol Sci* **59**:209–218 (2001).

45. P. Bernillon and F. Y. Bois, Statistical issues in toxicokinetic modeling: a Bayesian perspective. *Environ Health Perspect* **108**:883–893 (2000).

46. A. Roy, C. P. Weisel, P. J. Lioy, and P. G. Georgopoulos, A distributed parameter physiologically-based pharmacokinetic model for dermal and inhalation exposure to volatile organic compounds. *Risk Anal* **16**:147–160 (1996).

47. L. Fabrizi, G. W. Taylor, B. Canas, A. R. Boobis, and R. J. Edwards, Adduction of the chloroform metabolite phosgene to lysine residues of human histone H2B. *Chem Res Toxicol* **16**:266–275 (2003).

48. R. A. Corley, A. L. Mendrala, F. A. Smith, D. A. Staats, M. L. Gargas, R. B. Conolly, M. E. Andersen, and R. H. Reitz, Development of a physiologically based pharmacokinetic model for chloroform. *Toxicol Appl Pharmacol* **103**:512–527 (1990).

APPENDIX 43.1   CODE LISTING 1

This Matlab code is for the main chloroform PBPK case study, which simulates the PBPK model for different exposure scenarios.

```
1.   function res = PBPK_chloroform_case_study
2.   % A case study using a PBPK model for chloroform exposures
3.
4.   % Defining the individual
5.   indiv.age = 30;
6.   indiv.sex = 'M';
7.   indiv.BW = 70; % Body Weight (kg)
8.   indiv.BSA = 18000; % Body Surface Area (cm^2)
9.
10.  modelopts = []; % all defaults => one compartment skin model
11.
12.  % start, end exposure, end simulation
13.  events.timestages = [0 30 360]; % minutes
14.  events.airconcs = [0.1 0]; % micro.g/L
15.  events.contact_media = [1 0]; % dermal contact with water, and then air
16.  events.waterconcs = [150 150]; % micro.g/L
17.  events.timeincrements = [1 1]; % output increments
18.  events.bolus_ingestion = [0 0]; % amount in micro.g
19.  events.bolus_injection = [0 0]; % amount in micro.g
20.  events.rate_inhale = [5.8 5.8]; % L/min
```

```
21.  events.drinking_water_rate = [0 0]; % L/min
22.  events.infusion = [0 0]; % micro.g/min
23.  orig_events = events; % saving the events for subsequent reuse
24.
25.  config = PBPK_chloroform_config(indiv, modelopts);
26.  % Additional changes to the PBPK configuration can be made here
27.
28.  % Case study results
29.  events = orig_events; % inhalation and dermal exposures
30.  res.oneskin.inh_dermal = PBPK_voc_driver(indiv, config, events);
31.
32.  events = orig_events;
33.  events.airconcs = [0 0]; % no inhalation exposure; dermal only
34.  res.oneskin.dermal_only = PBPK_voc_driver(indiv, config, events);
35.
36.  % Distributed parameter skin model
37.  modelopts.DM.nskins = 100; % nodes for distributed skin model
38.  config = PBPK_chloroform_config(indiv, modelopts); % New configuration
39.
40.  events = orig_events; % inhalation and dermal exposures
41.  res.dpskin.inh_dermal = PBPK_voc_driver(indiv, config, events);
42.
43.  events = orig_events;
44.  events.airconcs = [0 0]; % no inhalation exposure; dermal only
45.  res.dpskin.dermal_only = PBPK_voc_driver(indiv, config, events);
46.
47.  events = orig_events;
48.  events.waterconcs = [0 0]; % no dermal exposure; inhalation only
49.  res.dpskin.inh_only = PBPK_voc_driver(indiv, config, events);
```

APPENDIX 43.2   CODE LISTING 2

This Matlab code is for driving the PBPK model simulation based on the model
and event configuration. The function is not chemical specific.

```
1.   function simout = PBPK_voc_driver(indiv, config, events, opt_amount_initial)
2.   % Main driver for the PBPK voc simulation
3.   % Global configuration variables are used to share information between this
4.   % function and derivative code, which is invoked indirectly via the
5.   % Ordinary Differential Equation (ODE) solver
6.   %
7.   % Model inputs:
8.   % indiv -- structure containing physiological parameters of individual
9.   % events -- structure containing information on the exposure events
10.  % config -- structure containing information on PBPK model configuration
11.  % opt_amount_initial -- (optional) initial amounts of chemical in tissues
```

```matlab
12.  % Model outputs:
13.  % simout -- structure containing time profiles of amounts and concentrations,
14.  % model and event configurations, as well as mass balance terms
15.
16.  global c E; % c is for Configuration and E is for Event Specification
17.  c = config;
18.
19.  if (nargin >= 4)
20.    amount = opt_amount_initial;
21.  else
22.    amount = zeros(c.SV.TotNum, 1); % All state variables set to zero
23.  end
24.
25.  nstages = length(events.timestages) - 1;
26.
27.  saved_amounts = [];
28.  saved_times = [];
29.  saved_mass_balances = [];
30.  saved_sv = [];
31.
32.  saved_amounts(end+1,:) = amount; % initial values
33.  saved_times(end+1,:) = events.timestages(1); % initial time
34.  saved_mass_balances(end+1,:) = 0; % initial values -- zeros
35.
36.  SVID = c.SV.ID; % state variable IDs
37.
38.  odeset('RelTol', 1e-8, 'AbsTol', 1e-12, 'MaxOrder', 5, 'BDF', 'on');
39.
40.  for istage=1:nstages
41.    stage_beg = events.timestages(istage);
42.    stage_end = events.timestages(istage+1);
43.    E = local_get_event(events, istage);
44.    tspan = [stage_beg:events.timeincrements(istage):stage_end];
45.    amount(SVID.GILumen) = amount(SVID.GILumen) + ...
46.      events.bolus_ingestion(istage) * c.F_bioavail_water;
47.    amount(SVID.Ingested) = amount(SVID.Ingested) + ...
48.      events.bolus_ingestion(istage);
49.    amount(SVID.Feces) = amount(SVID.Feces) + ...
50.      events.bolus_ingestion(istage) * (1 - c.F_bioavail_water);
51.    amount(SVID.Blood) = amount(SVID.Blood) + events.bolus_injection(istage);
52.
53.    [cur_time_new, amount_new] = ode15s('PBPK_voc_deriv',tspan, amount);
54.    saved_amounts = [saved_amounts(1:end-1,:); amount_new];
55.    mass_balances = local_check_mass_balance(c, amount_new);
56.    saved_mass_balances = [saved_mass_balances(1:end-1), ...
57.      mass_balances]; % save mass balances
58.    saved_times = [saved_times(1:end-1); cur_time_new];
59.
```

```
60.   % Save the intermediate state variables such as concentrations and rates
61.   saved_sv = saved_sv(1:end-1);
62.   for i=1:length(tspan)
63.     [dummy, optSV] = PBPK_voc_deriv(tspan(i), amount_new(i,:)');
64.     saved_sv = [saved_sv; optSV];
65.   end
66.
67.   amount = amount_new(end,:); % Set the state for the next stage
68.
69.   end
70.
71.   simout.amounts = saved_amounts; % amounts in tissues, etc
72.   simout.mass_balances = saved_mass_balances; % mass balance term vs time
73.   simout.times = saved_times; % times at which model outputs are saved
74.   simout.sv = saved_sv; % state variables of the model
75.   simout.PBPK_config = c;
76.   simout.event_config = events;
77.   simout.concs = [saved_sv.concs]; % main tissues, exhaled, and venous blood
78.
79.   function thisevent = local_get_event(events, istage)
80.   thisevent.Q.inh = events.rate_inhale(istage);
81.   thisevent.Q.DW = events.drinking_water_rate(istage);
82.   thisevent.C.inh = events.airconcs(istage);
83.   thisevent.C.water = events.waterconcs(istage);
84.   thisevent.dermal_contact_media = events.contact_media(istage);
85.   thisevent.Q.infusion = events.infusion(istage);
86.
87.   function res = local_check_mass_balance(c, orig_amount)
88.   SVID = c.SV.ID; % state variable IDs
89.   amount = orig_amount';
90.   % Body burden for all compartments (including stratum corneum)
91.   Body_burden = sum(amount(1:length(c.CompName),:));
92.   Body_burden = Body_burden + amount(SVID.GILumen,:) + amount(SVID.Blood,:);
93.   mass_difference = amount(SVID.Inhaled,:) - amount(SVID.Exhaled,:) . . .
94.     + amount(SVID.Ingested,:) - amount(SVID.Feces,:) . . .
95.     + amount(SVID.Dermal,:) - amount(SVID.Metabolized,:) - Body_burden;
96.
97.   input_dose = amount(SVID.Inhaled,:) + . . .
98.     amount(SVID.Ingested,:) + amount(SVID.Dermal,:);
99.   if (abs(input_dose) > eps)
100.    mass_balance_term = mass_difference ./ input_dose;
101.    if ( max(abs(mass_difference/input_dose)) > 0.0001 )
102.      error('*** mass balance error exceeds 0.01 % ***');
103.    end
104.  else
105.    mass_balance_term = zeros(1,size(amount,2));
106.  end
107.  res = mass_balance_term;
```

APPENDIX 43.3   CODE LISTING 3

This Matlab code is for the PBPK model configuration for chloroform. This produces a PBPK configuration object based on the individual's physiological parameters and other model configuration options.

```
1.   function c = PBPK_chloroform_config(indiv, modelconf)
2.   % Provides default configuration values for the PBPK_inhalation model
3.   % Structure c holds the complete PBPK Model Configuration
4.   % General parameters obtained from Roy et al., Risk Analysis 16(2)
5.   % Exception: Q_cardiac from Fisher et al., 1999
6.   %
7.   % Usage: x = PBPK_inhalation_default_chloroform_config(indiv, modelconf)
8.   % Function Inputs:
9.   %  indiv: a structure with the following fields
10.  %  age -- age of the individual
11.  %  sex -- sex of the individual ('M' or 'F')
12.  %  BW  -- body weight of the individual
13.  %  BSA -- body surface area
14.  % modelconf: a structure that defines model configuration with fields
15.  % DM.nskins -- specifies type of dermal model to use
16.  %  1  => one skin model
17.  %  2  => two skin model (Stratum Corneum (SC) + Viable Skin)
18.  %  >2 => distributed skin model with nskins nodes in SC
19.  % if this structure is not provided, the default is a one-skin model
20.  %
21.  % Outputs
22.  % c - configuration object for the PBPK model parameters
23.  %
24.
25.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
26.  % PBPK Model Structure Parameters (e.g. modeled compartments)
27.  c.IDNum.Fat  = 1;  c.CompName{1} = 'Fat';  % Fat
28.  c.IDNum.SP   = 2;  c.CompName{2} = 'SP';   % Slowly Perfused Tissue
29.  c.IDNum.RP   = 3;  c.CompName{3} = 'RP';   % Rapidly Perfused Tissue
30.  c.IDNum.Liv  = 4;  c.CompName{4} = 'Liv';  % Liver
31.  c.IDNum.Gut  = 5;  c.CompName{5} = 'Gut';  % Gut
32.  c.IDNum.Kid  = 6;  c.CompName{6} = 'Kid';  % Kidney
33.  c.IDNum.Skin = 7;  c.CompName{7} = 'Skin'; % Skin
34.
35.  ID=c.IDNum; % A temporary variable for compartment IDs
36.  c.N_Compartments = length(c.CompName);
37.  c.Density = 1.0; % Body and tissue density
38.
39.  % Set important variables to NaN. Helps in identify initialization errors
40.  nanArray=NaN(1,c.N_Compartments);%NaNs corresponding to each compartment
41.  c.PC.tissue = nanArray; %Blood partition coefficients for all compartments
42.
```

```
43.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
44.  % Chemical-independent physiological parameters (volume ratios, etc)
45.  c.AF.QC.M = 15.87; % Allometric factor (AF) for cardiac output (Males)
46.  c.AF.QC.F = 17.7; % AF (Females). Units: L/hour/kg.
47.
48.  % Percentage of Cardiac Flow Ratios
49.  c.QCR(ID.Fat)  =  0.05;
50.  c.QCR(ID.SP)   =  0.156;
51.  c.QCR(ID.RP)   =  0.26;
52.  c.QCR(ID.Liv)  =  0.07;
53.  c.QCR(ID.Gut)  =  0.18;
54.  c.QCR(ID.Kid)  =  0.25;
55.  c.QCR(ID.Skin) =  0.04;
56.  c.QCR(ID.SP)   =  c.QCR(ID.SP)+1-sum(c.QCR);%AdjustQCRforSPtosumto1
57.
58.  c.venous_blood_volume_fraction = 0.01;
59.
60.  c.VR(ID.Fat)      =  0.231;
61.  c.VR(ID.SP)       =  0.5105;
62.  c.VR(ID.RP)       =  0.0327;
63.  c.VR(ID.Liv)      =  0.0314;
64.  c.VR(ID.Gut)      =  0.017;
65.  c.VR(ID.Kid)      =  0.0044;
66.  c.VR(ID.Skin)     =  0.1;
67.
68.  % Adjust VR for SP to sum to 1
69.  c.VR(ID.SP)=c.VR(ID.SP)+1-sum(c.VR)-c.venous_blood_volume_fraction;
70.
71.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
72.  % Chemical Specific Information
73.  c.Notes = 'Chloroform PBPK Model for Inhalation Route';
74.  c.Compound = 'Chloroform';
75.  c.MolecularWeight = 119.4;
76.
77.  % Chemical Specific Partition Coefficients for Chloroform
78.  c.PC.blood_air = 7.43; % blood_air PC
79.  c.PC.tissue(ID.Fat)  = 37.7; % fat/blood PC
80.  c.PC.tissue(ID.SP)   =  1.62; % SP/blood PC
81.  c.PC.tissue(ID.RP)   =  2.3;
82.  c.PC.tissue(ID.Gut)  =  2.3;
83.  c.PC.tissue(ID.Liv)  =  2.3;
84.  c.PC.tissue(ID.Kid)  =  1.5;
85.
86.  c.PC.chem.air_water = 0.163; % Henry's Law Coefficient from USEPA
87.
88.  c.MMK.VMAX_C = 0.26167; % Allometric constant for Vmax
89.  c.MMK.KM_C = 0.448 * 1000; % Michaelis-Menten constant in liver (ug/L)
90.
```

```matlab
91.  c.Absorption.Gut = 1.0; % Gut-Lumen absorption rate constant (1/hr)
92.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
93.  % Updating the parameters based on the individual
94.  c.indiv = indiv;
95.  BW = c.indiv.BW; % simpler variable for some formulae
96.
97.  c.V_TISSUE = BW * c.Density * c.VR; % Tissue volumes from volume fractions
98.  c.venous_blood_volume = c.venous_blood_volume_fraction * BW * c.Density;
99.  c.Q_cardiac = c.AF.QC.(c.indiv.sex)/60 *(BW)^0.75; % Units: L/min -- Fisher
100. c.VMAX = zeros(1,c.N_Compartments); % VMAX initialization (Michelis Menten)
101. c.VMAX(ID.Liv) = c.MMK.VMAX_C * 1000 * BW^0.7; % Units: ug/min
102. c.KM = ones(1,c.N_Compartments) * (-Inf);
103. c.KM(ID.Liv) = c.MMK.KM_C;
104.
105. c.F_bioavail_water = 1; % Oral bioavailability from water
106. c.QC = c.QCR * c.Q_cardiac; % Blood flow rates to each compartment
107.
108. if (nargin < 2)
109.     modelconf = []; % default option
110. end
111.
112. c = PBPK_add_dermal_model(c, modelconf); % Add the dermal model
113.
114. % State Variables in the Model
115. c.SV.ID = c.IDNum; % State Variables ID
116.
117. c.SV.Names = c.CompName;
118. c.SV.Names{end+1} = 'Blood'; % Amount in Blood
119. c.SV.Names{end+1} = 'GILumen';
120. c.SV.Names{end+1} = 'Metabolized';
121. c.SV.Names{end+1} = 'Inhaled';
122. c.SV.Names{end+1} = 'Exhaled';
123. c.SV.Names{end+1} = 'Feces';
124. c.SV.Names{end+1} = 'Dermal';
125. c.SV.Names{end+1} = 'Ingested';
126. c.SV.Names{end+1} = 'Dermal_VS';
127. c.SV.Names{end+1} = 'AUCL';
128.
129. n_comp = c.N_Compartments + c.DM.nskin_comps;
130. for i=length(c.CompName)+1:length(c.SV.Names)
131.     thisSVName = c.SV.Names{i};
132.     c.SV.ID.(thisSVName) = n_comp + 1;
133.     n_comp = n_comp + 1;
134. end
135.
136. c.SV.TotNum = n_comp; % Total Number of State Variables in this PBPK Model
137.
138. return;
```

APPENDIX 43.4   CODE LISTING 4

This Matlab code is for updating the dermal portion of the PBPK model configuration for chloroform, based on the model configuration options.

```matlab
1.    function c = PBPK_add_dermal_model(c_orig, modelconf)
2.    % Updates the PBPK model configuration with dermal parameters
3.    %
4.    % Function Inputs:
5.    % c_orig: the configuration of the PBPK model prior to model update
6.    % modelconf: a structure that defines model configuration with fields
7.    % DM.nskins - specifies type of dermal model to use
8.    %   1  => one skin model
9.    %   2  => two skin model (Stratum Corneum (SC) + Viable Skin)
10.   %   >2 => distributed skin model with nskins nodes in SC
11.   % if the DM structure is not provided, or if there is no DM.nskins
12.   % field, the default is a one-skin model
13.   %
14.   % Function Outputs:
15.   %   c: the updated PBPK model configuration
16.   %
17.   % Function side effects:
18.   %   Depending on the model configuration, the skin compartment in the
19.   %   default PBPK model may be replaced by SC, VS, and other compartments
20.
21.   nskins = 1; % default number of skin compartments
22.   if (isfield(modelconf, 'DM') & isfield(modelconf.DM, 'nskins'))
23.     nskins = modelconf.DM.nskins;
24.     if (nskins <= 0)
25.       error(['Invalid number of skins in Dermal Model: ' num2str(nskins)]);
26.     end
27.   end
28.
29.   c_derm.distributed = 0; % default is non-distributed model
30.
31.   switch(nskins)
32.   case (1)
33.     c_derm.n_sc = 0;        % zero for default one skin case
34.     c_derm.nn_sc = 0;       % number of Stratum Corneum nodes
35.     c_derm.nskins = 1;
36.   case (2)
37.     c_derm.n_sc = 1;        % zero for default one skin case
38.     c_derm.nn_sc = 0;       % number of Stratum Corneum nodes
39.     c_derm.nskins = 2;
40.   otherwise
41.     c_derm.n_sc = 1;        % zero for default one skin case
42.     c_derm.nn_sc = nskins; % number of Stratum Corneum nodes
43.     c_derm.nskins = nskins;
```

```
44.    c_derm.distributed = 1;
45.  end
46.
47.  c_derm.nskin_comps = c_derm.n_sc + c_derm.nn_sc;
48.
49.  c = c_orig;
50.  c.DM = c_derm; % c.DM stands for the dermal model configuration
51.  ID = c.IDNum;
52.
53.  c.PC.tissue(ID.Skin) = 1.6; % Skin-blood partition coefficient
54.  c.K.skin_water = 1.48E-4; % Skin permeability [cm/min]
55.
56.  c.PCMC.skin_air = c.PC.tissue(ID.Skin) * c.PC.blood_air;
57.  c.PCMC.skin_water = c.PCMC.skin_air * c.PC.chem.air_water;
58.  c.K.skin_air = c.K.skin_water * c.PCMC.skin_air/c.PCMC.skin_water;
59.
60.  c.DM.L_sc = 0.001; % thickness of stratum corneum [cm]
61.
62.  if (nskins > 1)
63.   c = local_update_dermal_parameters(c, nskins);
64.   c.K.skin_air = c.K.sc_air; % SC replaces skin
65.   c.K.skin_water = c.K.sc_water;
66.  end
67.
68.  return;
69.
70.
71.  function c = local_update_dermal_parameters(c_orig, nskins)
72.  c = c_orig;
73.  ID = c.IDNum;
74.
75.  % Two-skin compartment
76.  c.DM.V_sc = c.DM.L_sc * c.indiv.BSA;
77.  c.DM.Vskin = c.V_TISSUE(ID.Skin) * 1000; % in cm^3
78.
79.  % Partition coefficients between multiple chemicals (non-blood)
80.  c.PC.chem.octanol_water = 90;
81.
82.  % Partition coefficients between multiple compartments (non-blood)
83.  c.PCMC.sc_water = (c.PC.chem.octanol_water)^0.71;
84.  c.PCMC.sc_air = c.PCMC.sc_water/c.PC.chem.air_water;
85.  c.PCMC.vs_water = (c.PCMC.skin_water * c.DM.Vskin - ...
86.          c.PCMC.sc_water * c.DM.V_sc)/(c.DM.Vskin - c.DM.V_sc);
87.  c.PCMC.vs_air = c.PCMC.vs_water/c.PC.chem.air_water;
88.  c.PCMC.vs_blood = c.PCMC.vs_air/c.PC.blood_air;
89.  c.PCMC.vs_sc = c.PCMC.vs_air/c.PCMC.sc_air;
90.
91.  %%% Calculate Cleek and Bunge's B parameter %%%
```

```
 92.  c.DM.BB = c.K.skin_water * sqrt(c.MolecularWeight)/2.6;
 93.  c.K.sc_water = c.K.skin_water * (c.DM.BB + 1);
 94.  c.K.sc_air = c.K.sc_water * c.PCMC.sc_air/c.PCMC.sc_water;
 95.  c.K.vs_sc = c.K.sc_water/(c.PCMC.sc_water * c.DM.BB);
 96.
 97.  % Assign compartment IDs
 98.  c.IDNum.VS = ID.Skin;
 99.  c.IDNum = rmfield(c.IDNum, 'Skin');
100.  c.CompName{c.IDNum.VS} = 'VS';
101.  c.IDNum.SC = c.IDNum.VS + 1;
102.  c.CompName{c.IDNum.SC} = 'SC';
103.
104.  c.PC.tissue(c.IDNum.VS) = c.PCMC.vs_blood;
105.
106.  if (c.DM.distributed == 1)
107.   c.DM.D_sc = c.K.sc_water/(c.PCMC.sc_water) * c.DM.L_sc;
108.   c.IDNum.SC1 = c.IDNum.SC + 1;
109.   c.CompName{c.IDNum.SC1} = 'InnerMost SC Node';
110.  end
```

APPENDIX 43.5   CODE LISTING 5

This Matlab code is for calculating the derivatives used in the PBPK model.

```
 1.  function [d_amount, optSV] = PBPK_voc_deriv(t, amount)
 2.  % Derivative function defined in a manner the ODE solver in Matlab expects.
 3.  % Inputs:
 4.  %  t -- current time of the simulation (time variable in the ODE system)
 5.  %  amount -- amount of chemical in each tissue (state variables)
 6.  % Outputs: d_amount -- derivatives
 7.  %  optSV -- optional state variables such as intermediate concentrations
 8.  % Global: Configuration variable of the PBPK model
 9.
10.  global c; % Global PBPK model configuration structure
11.  global E; % Global activity event details
12.
13.  mc.air = E.C.inh; % media concentration: air; inhalation concentration
14.  rate.inh = E.Q.inh; % inhalation rate
15.  mc.water = E.C.water; % media concentration: water; ingestion concentration
16.  rate.dw = E.Q.DW; % ingestion rate -- drinking water
17.  rate.infusion = E.Q.infusion;
18.
19.  ID = c.IDNum; % Compartment IDs by name
20.  SVID = c.SV.ID; % State Variables ID
21.
22.  % Tissue Concentrations
23.  C_TISSUE = amount(1:length(c.V_TISSUE))' ./ c.V_TISSUE;
24.
```

```
25.  % Venous Blood Concentration
26.  CV_BLOOD = amount(1:length(c.V_TISSUE))' ./ (c.V_TISSUE .* c.PC.tissue);
27.
28.  CV_mixed = (CV_BLOOD*c.QC')/c.Q_cardiac;
29.  % Reroute the venous flow rate from Gut through Liver
30.  CV_mixed = CV_mixed + (CV_BLOOD(ID.Liv) - CV_BLOOD(ID.Gut)) * ...
31.              c.QC(ID.Gut)/c.Q_cardiac;
32.  venous_blood_concentration = amount(SVID.Blood)/c.venous_blood_volume;
33.
34.  C_arterial=(venous_blood_concentration*c.Q_cardiac+mc.air*rate.inh) ...
35.              / (c.Q_cardiac + rate.inh/c.PC.blood_air);
36.  mc.exh = C_arterial/c.PC.blood_air; % Exhaled concentration
37.
38.  % Rate of metabolism of chemical in each compartment
39.  rate_metabolism = c.VMAX .* CV_BLOOD ./ (c.KM + CV_BLOOD);
40.  d_amount = (c.QC .* (C_arterial - CV_BLOOD) - rate_metabolism);
41.  % Add the flow of venous blood from Gut to Liver to Venous Blood
42.  d_amount(ID.Liv) = d_amount(ID.Liv) + c.QC(ID.Gut)* ...
43.              (CV_BLOOD(ID.Gut) - CV_BLOOD(ID.Liv));
44.
45.  d_amount(SVID.Blood)=c.Q_cardiac*(CV_mixed-venous_blood_concentration)+...
46.              rate.infusion;
47.  d_amount(SVID.Metabolized) = sum(rate_metabolism);
48.  d_amount(SVID.AUCL) = C_TISSUE(ID.Liv);
49.  d_amount(SVID.Inhaled) = rate.inh * mc.air;
50.  d_amount(SVID.Exhaled) = rate.inh * mc.exh;
51.  d_amount(SVID.Ingested) = rate.dw*mc.water;
52.  rate_gut_absorption = c.Absorption.Gut * amount(SVID.GILumen);
53.  d_amount(SVID.GILumen)=d_amount(SVID.Ingested)*c.F_bioavail_water-...
54.              rate_gut_absorption;
55.  d_amount(ID.Gut) = d_amount(ID.Gut) + rate_gut_absorption;
56.  d_amount(SVID.Feces)=d_amount(SVID.Ingested)*(1-c.F_bioavail_water);
57.
58.  % Update the derivative for additional processes
59.  d_amount = PBPK_add_dermal_derivative(amount, d_amount, c, E);
60.
61.  d_amount = d_amount';
62.
63.  optSV.concs.tissue = C_TISSUE; % All tissues
64.  optSV.concs.exh = mc.exh; % exhaled breath concentrations
65.  optSV.concs.venous_blood = venous_blood_concentration;
66.  optSV.rate.metabolism = rate_metabolism;
```

APPENDIX 43.6   CODE LISTING 6

This Matlab code is for updating the PBPK model derivatives based on the dermal model options.

```
1.  function d_amount=PBPK_add_dermal_derivative(amount,d_amount_orig,c,E)
2.  % Updates to the basic PBPK model derivative structure with dermal derivative
3.  % Inputs are same as the generic PBPK_update_derivative function
4.  % Output: d_amount -- derivative vector updated with the dermal model equations
5.
6.  d_amount = d_amount_orig;
7.
8.  if (E.dermal_contact_media == 0)
9.    mc.dermal = E.C.inh; % dermal/air contact
10.   k_skin_media = c.K.skin_air; % Permeability: skin/air
11.   pc_skin_media = c.PCMC.skin_air; % Partition coefficient: skin/air
12.   if (c.DM.nskins > 1)
13.     pc_skin_media = c.PCMC.sc_air; % Partition coefficient: sc/air
14.   end
15. else
16.   mc.dermal = E.C.water; % dermal/water contact
17.   k_skin_media = c.K.skin_water; % Permeability: skin/water
18.   pc_skin_media = c.PCMC.skin_water; % Partition coefficient: skin/water
19.   if (c.DM.nskins > 1)
20.     pc_skin_media = c.PCMC.sc_water; % Partition coefficient: sc/water
21.   end
22. end
23.
24. SVID = c.SV.ID; % State variable IDs by name
25. % Tissue Concentrations
26. C_TISSUE = amount(1:length(c.V_TISSUE))' ./ c.V_TISSUE;
27.
28. switch(c.DM.nskins)
29.   case (1) % One compartment Skin = VS
30.   C_skin = amount(SVID.Skin)/c.V_TISSUE(SVID.Skin); % Concentration in skin
31.   rate_dermal_abs = (k_skin_media * c.indiv.BSA)/1000 *...
32.     (mc.dermal - C_skin/pc_skin_media);
33.   d_amount(SVID.Skin) = d_amount(SVID.Skin) + rate_dermal_abs;
34.   d_amount(SVID.Dermal) = rate_dermal_abs;
35.   d_amount(SVID.Dermal_VS) = rate_dermal_abs;;
36.   case (2) % Two comparments: SC and VS
37.   C_sc = amount(SVID.SC)/c.DM.V_sc; % Concentration in SC
38.   C_vs = amount(SVID.VS)/c.V_TISSUE(SVID.VS); % Concentration in viable skin
39.   rate_dermal_abs = (k_skin_media * c.indiv.BSA)/1000 *...
40.     (mc.dermal - C_sc/pc_skin_media);
41.   rate_vs_abs = (c.K.vs_sc * c.indiv.BSA)/1000 * (C_sc - C_vs/c.PCMC.vs_sc);
42.   d_amount(SVID.SC) = rate_dermal_abs - rate_vs_abs;
43.   d_amount(SVID.VS) = d_amount(SVID.VS) + rate_vs_abs;
44.   d_amount(SVID.Dermal) = rate_dermal_abs;
45.   d_amount(SVID.Dermal_VS) = rate_vs_abs;
46.   otherwise % Distributed PDE model
47.   % Formualte the 1-D PDE in terms of finite differences
48.   nn = c.DM.nn_sc; % number of nodes
```

```
49.   dl = c.DM.L_sc/(nn-1); % length of each node
50.
51.   % Boundary Conditions
52.   C_sc_in = C_TISSUE(SVID.VS)/c.PCMC.vs_sc; % Innermost SC: SC1
53.   C_sc_out = mc.dermal * pc_skin_media; % Outermost SC layer
54.   ID_sc_out = SVID.SC1 + nn - 1; % ID for outermost SC
55.   C_sc = [C_sc_in; amount(SVID.SC1+1:ID_sc_out-1); C_sc_out];
56.
57.   % Central Differnce Forumation across nodes of SC
58.   lnodes = C_sc(1:nn-2); % array containing left side nodes
59.   mnodes = C_sc(2:nn-1); % middle nodes
60.   rnodes = C_sc(3:nn); % right side nodes
61.   d_C_sc_nodes = c.DM.D_sc/dl^2 * (lnodes - 2*mnodes + rnodes);
62.
63.   % Dermal absorption rates on both sides of SC
64.   bsa = c.indiv.BSA; % cm^2
65.   rate_dermal_abs = c.DM.D_sc/dl * bsa * (C_sc(nn)-C_sc(nn-1))/1000;
66.   rate_vs_abs = c.DM.D_sc/dl * bsa * (C_sc(2) - C_sc(1))/1000; % innermost
67.   d_amount(SVID.VS) = d_amount(SVID.VS) + rate_vs_abs;
68.   d_amount(SVID.SC)=rate_dermal_abs-rate_vs_abs;%entireSCconcentration
69.   d_amount(SVID.SC1:SVID.SC1+nn-1) = [0 d_C_sc_nodes' 0]; % middle nodes
70.   d_amount(SVID.Dermal) = rate_dermal_abs;
71.   d_amount(SVID.Dermal_VS) = rate_vs_abs;
72. end
```

Modeling of Metabolite Pharmacokinetics in a Large Pharmacokinetic Data Set: An Application

VALÉRIE COSSON, KARIN JORGA, and ELIANE FUSEAU

44.1   INTRODUCTION

When a drug or a prodrug is metabolized to one or more active metabolites, not only the exposure to the parent drug but also the exposure to the active metabolites contribute to the safety and efficacy of that drug/prodrug (1–7). Prodrugs represent an aspect of the biotransformation of parent drug, where only the metabolite is active. Often, the prodrug is not subjected to intense pharmacokinetic (PK) modeling since its concentration declines rapidly; the absorption parameters of the active moiety encompass the transformation of the parent drug to its active form and absorption of the parent.

The blood or plasma concentrations of the parent drug and/or its active metabolites (systemic exposure) may provide an important link between drug dose (exposure) and desirable and/or undesirable drug effects (8). For this reason, the modeling of parent drug and metabolite pharmacokinetics, coupled with pharmacodynamic (PD) measurements, offers an essential development tool for prediction and simulation.

The simultaneous modeling of parent drug and metabolite allows the evaluation of the impact of organ impairment or of the effects of drug–drug interactions (9–13). The high incidence of adverse events seen in patients with end stage renal disease may, for some drugs, be explained in part by the accumulation of active drug metabolites (1). Any interaction at the site of drug metabolizing enzymes can modify the overall activity of the compound. It is often informative to have the prediction of metabolite concentrations when performing PK/PD modeling (2–7). Delay between the concentration of the parent compound and drug response curve (hysteresis) can be the result of metabolism when the metabolite is more effective than the parent drug.

Unless the metabolite has been administered alone (in order to estimate its volume of distribution (14)) or the fraction of parent drug converted to metabolite is known, the modeling of the parent drug and its metabolite requires simplification so that metabolite parameters can be estimated. This is because the rate of conversion of the parent to the metabolite and the distribution volume of the metabolite are structurally not simultaneously identifiable. The ratio of the rate of conversion to the metabolite volume, which is globally identifiable, and apparent elimination rate constant for the metabolite can be estimated and do not constitute an identifiability problem (15–21). The number of metabolites for which parameters can be estimated is not limited. For example, the parent compound, active metabolites, and its conjugate have been modeled together for irinotecan (15). More complicated models can be employed to investigate autoinducible transformations (9, 16, 22).

44.2   THE NELFINAVIR EXAMPLE

Intra- and interindividual variations in protease inhibitor drug exposure can influence the safety and effectiveness of anti-HIV therapy. In the example presented here, in addition to the description of the pharmacokinetics of nelfinavir and its metabolite, it was possible to evaluate the impact of the coadministration of ritonavir.

Nelfinavir is the only marketed HIV protease inhibitor that is converted into an active metabolite at plasma levels, which are significant enough to contribute to the overall antiviral activity (23).

Nelfinavir distributes largely into tissues and is highly bound to plasma proteins (>98%). The apparent volume of distribution is 2–7 L/kg. Nelfinavir is metabolized in the liver by at least four different cytochrome P450 (CYP) isoenzymes including CYP 3A4, CYP 2C9, CYP 2C19, and CYP 2D6, with CYP 2C19 catalyzing roughly 50% of nelfinavir clearance in normal metabolizers (23–25). CYP 2C19 mediates the formation of the primary metabolite M8 (nelfinavir hydroxy-*t*-butylamide), which has activity comparable to the parent drug. M8 is subsequently metabolized by CYP 3A4. The majority of nelfinavir and its metabolites (87%) are eliminated in the feces. Urinary excretion accounts for only 1–2%, most of which is unchanged nelfinavir (26). Nelfinavir induces its own metabolism; plasma concentrations decline approximately 40–50% and are stable after 6 days (26, 27).

Ritonavir (RTV) is also an inhibitor of HIV proteases, approved for use in combination with nucleoside analog, for the treatment of HIV-1 infected adults, adolescents, and children. It is a potent CYP 3A4 inhibitor and is used at low doses to elevate plasma concentrations of other protease inhibitors being primarily metabolized by CYP 3A4. In combination with saquinavir, this type of interaction has proved favourable (28). The combination with nelfinavir showed much smaller effects on nelfinavir levels, but it appears to change in normal metabolizers the M8/nelfinavir concentration ratio from 0.3 to 0.6. In poor CYP 2C19 metabolizers (~3–5% of Caucasians and African-Americans, ~12–20% of Asians), ritonavir addition is not expected to have such an effect on the nelfinavir/M8 ratio (29). In addition, ritonavir induces CYP isoenzymes, so that the full effect of the nelfinavir–ritonavir drug–drug interaction is considered stable after a treatment duration of 10–14 days (30).

44.2.1   Methods

44.2.1.1   *Study Design and Data*

The study was a randomized, stratified, open label, two-arm parallel Phase 4 trial designed to explore the utility of pharmacological parameters as predictors of antiviral response to nelfinavir-containing regimens in patients pretreated with protease inhibitors-sparing regimens.

The study population included male or female patients aged 18 years or older with HIV-1 infection who have previously failed only one antiretroviral regimen consisting of nucleoside reverse transcriptase inhibitors (NRTIs), nonnucleoside reverse transcriptase inhibitors (NNRTIs), or at most one protease inhibitor (PI) and had no evidence of PI resistance (genotypic) at screening or had plasma HIV-1 RNA greater or equal to 1000 copies/mL at least once in the three months before screening and HIV-1 RNA greater or equal to 1000 copies/mL at screening. Participants in the study were randomized to receive nelfinavir at 1250 mg BID or nelfinavir at 1250 mg BID in combination with ritonavir at 200 mg BID in a 1:1 ratio. Both treatments were administered for 48 weeks. In the treatment arm initially receiving nelfinavir as the only PI, ritonavir was added at week 12 if the HIV-1 RNA had not been reduced to <400 copies/mL or reduced by 2 logs compared to baseline. All patients received concomitant therapy with two NRTIs selected based on clinical judgment and the results of the genotypic and phenotypic resistance test performed at screening. Randomization was stratified according to the patient's individual sensitivity to the NRTIs prescribed to ensure that the treatment groups were balanced.

Eighty-three patients provided concentration–time data. Exclusion criteria included elevated transaminases, bilirubin, or serum creatinine; decreased neutrophils, hemoglobin, or platelet counts; malabsorption syndrome; opportunistic infections; alcohol, narcotics, barbiturates, cocaine, or other CNS-active substance abuse; hypersensitivity to any of the protocol mandated drugs; patients of childbearing potential who were unwilling to use an effective method of contraception; and concomitant medications interfering with human cytochrome P450 system.

Thirty-six patients received nelfinavir alone, 39 received nelfinavir plus ritonavir, and 8 patients received nelfinavir alone for 12 weeks and nelfinavir plus ritonavir thereafter. In the data set, there were 611 nelfinavir concentrations, among those 18 BLQ concentrations were replaced with half of lower quantification limit, and 611 M8 concentrations, among those 16 BLQ concentrations were replaced with half of quantification limit.

Summary of demographic and baseline characteristics are reported in Table 44.1. There were 18 female patients and 65 male patients. Forty-three patients were Caucasian, two patients were Black, 36 patients were Hispanic, and two patients belonged to another ethnic group. The median age was 37 years. The median of body weight was 68.5 kg and the median of height was 170 cm.

Twenty-three patients had abnormal liver function. The median alkaline phosphatase was 97 U/L, alanine aminotransferase was 33 U/L, aspartate aminotransferase was 28 U/L, total bilirubin was 11 μmol/L, gamma glutamyltransferase was 35 U/L, and total protein was 79 g/L. Within the patient population, the treatment groups (i.e., nelfinavir alone or nelfinavir plus ritonavir) were homogeneous.

TABLE 44.1   Patient Demographic and Baseline Characteristics

Characteristics	NFV+RTV	NFV Alone	NFV / NFV+RTV	All
Patients	39	36	8	83
Male	35	24	6	65
Female	4	12	2	18
Liver function class				
Normal	29	25	6	60
Abnormal	10	11	2	23
RNA class				
>75,000 copies/mL	4	9	2	15
10,000–75,000 copies/mL	19	15	1	35
<10,000 copies/mL	16	12	5	33
CD4 class				
<100 cells/mL	1	0	0	1
100–300 cells/mL	11	13	1	25
>300 cells/mL	27	23	7	57
Median age in years (range)	38.0 (25.0–69.0)	37.5 (22.0–55.0)	29.0 (24.0–44.0)	37.0 (22.0–69.0)
Median weight in kg (range)	68.5 (55.2–108.0)	65.4 (41.5–96.0)	73.0 (58.1–130.0)	68.5 (41.5–130.0)
Median height in cm (range)	170 (152–185)	170 (149–191)	173 (156–182)	170 (149–191)
Median body mass index in kg/m^2 (range)	23.7 (20.1–34.1)	23.1 (17.4–31.0)	24.0 (20.1–52.1)	23.5 (17.4–52.1)
Median lean body weight in kg (range)	54.7 (42.2–71.7)	51.7 (32.9–68.6)	55.8 (38.9–66.8)	53.8 (32.9–71.7)
Median body surface area in m^2 (range)	1.80 (1.59–2.25)	1.73 (1.32–2.14)	1.91 (1.63–2.23)	1.8 (1.32–2.25)
Median ALP in U/L (range)	99 (53–272)	100 (25–618)	95 (66–100)	97 (25–618)
Median AST in U/L (range)	29 (17–242)	27 (15–88)	26 (16–108)	28 (15–242)
Median ALT in U/L (range)	35 (14–408)	32 (14–121)	26 (11–135)	33 (11–408)
Median GGT in U/L (range)	35 (9–218)	35 (9–443)	40 (13–117)	35 (9–443)
Median total bilirubin in μmol/L (range)	11 (5–25)	11 (5–27)	9 (3–16)	11 (3–27)
Median total protein in g/L (range)	79 (69–93)	78 (66–93)	80 (75–82)	79 (66–93)
Median RNA copies/mL (range)	14,000 (757–750,001)	37,589 (2,202–489,000)	3,507 (2,046–257,689)	18,200 (757–750,001)
Median CD4 count in cell/mL (range)	403 (56–988)	352 (105–746)	402 (172–872)	368 (56–988)

Disease progression markers were measured in patients. The median CD4 count was 368 and viral load (i.e., HIV-1 RNA) was 18,200 copies/mL. The nelfinavir plus ritonavir group had a larger CD4 count and a lower viral load at baseline than the nelfinavir alone group, although the clinical significance of this finding was unclear.

44.2.1.2 Sampling Procedure and Analytical Methods

Patients had blood samples drawn for analysis of nelfinavir and M8 at weeks 2, 8, 12, and 48 (or early termination). Each individual provided two blood samples per visit; a predose through sample and a postdose sample. The patients presented to the clinic having fasted for at least 3 hours and having taken the previous dose of nelfinavir (and ritonavir) according to the regular schedule. The patients were provided a meal or snack of at least 300 kcal, which was to be eaten within 30 minutes before dosing. After completion of the meal and immediately before dosing, the predose blood sample was drawn. The patients took the dose of study medication and provided the second blood sample collected between 2 and 6 hours after the last dosing. All the concentrations were considered to be at steady state.

Nelfinavir and M8 were determined in plasma using a validated LC/MS/MS; the lower limits of quantification were 4.0 ng/mL and 1.0 ng/mL, respectively.

44.2.1.3 Population Pharmacokinetic Model Development

Based on prior information collected in a patients' study (31) and on the data collected, a one-compartment model with a first-order absorption with or without a lag time was tested for the choice of the structural model for nelfinavir. As an extension of the parent drug model, a one-compartment model with first-order absorption and elimination was tested for the metabolite. The structure of the PK model used is presented in Figure 44.1.

In this model, the clearance of nelfinavir is partitioned between the formation route of M8 via CYP 2C19 and all other routes of elimination of nelfinavir via CYP 2D6, 2C9, and 3A4. The metabolic clearance of nelfinavir to M8 and the distribution volume of M8 are not identifiable separately. The actual parameter estimated is the microrate constant `K23`, equal to the ratio of metabolic clearance of nelfinavir to M8 to the distribution volume of M8.
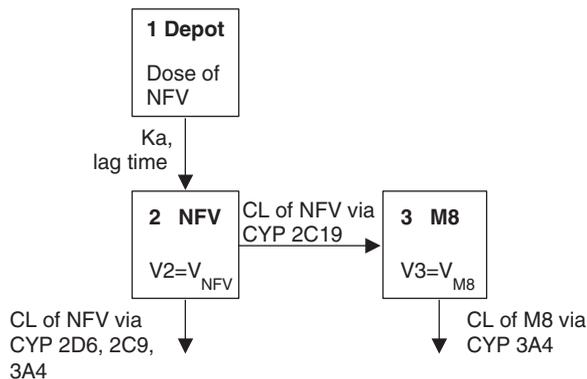
FIGURE 44.1   Schematic representation of the structural pharmacokinetic model.

Data were analyzed using the nonlinear mixed effects model software program NONMEM (Version 5 level 1.1 double precision (32)). Nelfinavir and M8 were fitted simultaneously. The molecular weight of nelfinavir and M8 is comparable with a ratio of M8 to nelfinavir of 1.028. Therefore, the concentrations were not corrected and are expressed in nanogram per milliliter.

Various writing of the control file could be used. With the coding 1, subroutine ADVAN 5, which implements a user-defined general linear model, was used. The model was parameterized in terms of clearance and volume with the following parameters to estimate:

CL, the clearance of NFV

V2, the volume of distribution of NFV

KA, the absorption rate constant for NFV

ALAG1, the lag time of NFV absorption

K23, the rate constant of formation of M8

CLM, the clearance of M8

The links between the clearances and volume and the microconstants were coded as follows:

```
K12 = KA
K20 = (CL/V2) − K23
K30 = CLM/V3
```

The distribution volume of M8 (V3) was fixed to 1 L and the estimated parameter was an apparent metabolic clearance of nelfinavir to M8.

The same writing was used but with the distribution volume of M8 fixed to be equal to the distribution volume of nelfinavir with the coding 2.

With the coding 3, the fraction of nelfinavir clearance for the formation of M8 (i.e., FMET) was modeled instead of K23. Since the distribution volume of M8 was fixed to 1 L, the fraction is interpreted as the ratio of the fraction nelfinavir converted to M8 and the distribution volume of M8. The links between the clearances and volume and the microconstants were set as follows:

```
K12 = KA
K20 = (1-FMET)*CL/V2
K23 = FMET*CL/V2
K30 = CLM/V3
```

With the coding 4, subroutine ADVAN 6, which implements a general nonlinear model user-defined with differential equations describing the process, was used. The model was parameterized in clearance and volume. The distribution volume of M8 was fixed to 1 L. The differential equations used are presented below:

$$\frac{dA(1)}{dt} = -KA \times A(1) \tag{44.1}$$

$$\frac{dA(2)}{dt} = KA \times A(1) - \frac{CL}{V2} \times A(2) \tag{44.2}$$

$$\frac{dA(3)}{dt} = K23 \times A(2) - \frac{CLM}{V3} \times A(3) \tag{44.3}$$

The complete NONMEM control files are provided in Appendixes 44.1–44.4. Model development was performed with the coding 1 only. Once the final model was obtained with the definitive covariate effect model on parameters and error models, the alternative writings (coding 2–4, Appendixes 44.2–44.4) were evaluated with the same covariate effect and error models as the final model with the coding 1 (see Appendix 44.1).

The influence of ritonavir on the pharmacokinetics of nelfinavir was investigated under the assumption that steady-state conditions for ritonavir have been reached. Thus, ritonavir-induced inhibition of nelfinavir and M8 metabolism via CYP 3A4 was assumed.

A proportional error, a constant additive error, and a combination of both error models were evaluated for the residual error model. Between-subject random effects were explored on the clearance of parent drug and metabolite, the volume of distribution of the parent drug, and the absorption rate constant. An exponential model was preferred. Interoccasion random effects were explored on the clearance of the parent drug and of the metabolite, the volume of distribution of the parent drug, and the absorption rate constant. An exponential model was preferred. The joint distribution of the between-subject random effect, the interoccasion random effects, and the residual error were assumed normal with mean 0 and variance–covariance matrices Ω for the between-subject and interoccasion random effects, and Σ for the residual error to be estimated. The FO method was used for the estimation of the parameters.

The effect of the following covariates was investigated on the disposition parameter of nelfinavir, M8, and ritonavir, for which a between-subject variance was estimated:

- Demographic data (body size, age, sex, ethnicity)
- HIV-1 infection markers (HIV-1 RNA and CD4 count, either as continuous or categorical variable)
- Blood biochemistry (alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, total bilirubin, gamma glutamyltransferase, total protein)

The selection of the structural PK model and residual error models was based on the goodness-of-fit plots and on the difference in NONMEM objective function (approximately $-2 \times$ log likelihood) between hierarchical models (i.e., the likelihood ratio test). This difference is asymptomatically χ^2 distributed with a degree of freedom equal to the number of additional parameters of the full compared to the reduced model. A p-value of 0.05 was chosen for one additional parameter, corresponding to a difference in the objective function of 3.84. Potential covariates were selected by univariate analysis, testing the addition of each covariate on each of the relevant PK parameters. When a set of covariates, identified by the

univariate selection, was found to be significant predictors of a parameter based on the likelihood ratio test, all were included in a full model. Backward deletion of these covariates, one at a time, was performed with a significance level of 0.005, that is, a drop in the objective function of at least 7.8. Thus, the final irreducible model was identified.

Acceptable population models resulted in successful minimization, with at least three significant digits for any parameter, a successful estimation of the covariance, and the absolute value of last iteration gradients greater than 0.001 but smaller than 100. Confidence intervals of structural parameters should not include value zero; correlation between any two structural parameters should never be greater than 0.95. Acceptable models should not lead to trends in the distribution of weighted residuals versus model predictions and versus independent variable. They should not be oversensitive to initial estimates nor lead to differences between the population parameters and the corresponding medians of individual POSTHOC parameters. The predictions versus observations data should be evenly distributed around the unit line. If constraints were applied on parameters, no final estimate should be equal to one of the boundaries.

The final model (coding 1) was compliant with the above model-acceptance criteria since the run finished successfully with more than three significant digits and the covariance, the 95% confidence intervals of all the parameters did not include zero, none of the correlation between the structural parameters was above 0.95, and the weighted residuals versus model predictions and versus independent variable data were evenly distributed around the zero line. However, slight trends toward overprediction were seen in the plot of predictions versus observations; possible explanations are given in Section 44.2.3.

44.2.2 Results

44.2.2.1 Distribution Concentrations of Nelfinavir and Its Metabolites
Concentration–time data of nelfinavir by treatment group are displayed as scatterplots in Figure 44.2. Concentration–time data of M8 by treatment group are displayed as scatterplots in Figure 44.3. The disposition also suggests a mono-exponential decay. The shapes of the profile for nelfinavir and M8 are quite similar.

44.2.2.2 Nelfinavir Alone
The nelfinavir data were fitted alone first. The between-subject variances were estimated on the clearance and the volume of distribution, and the interoccasion variance was estimated on the clearance. The residual error was modeled with a proportional error model. There was no statistically significant effect of ritonavir on clearance of nelfinavir. The population parameters of base model for nelfinavir alone are presented in Table 44.2. The apparent half-life of 7.0 h was driven by the absorption rate constant characterizing a flip-flop phenomenon. The goodness-of-fit plots are presented in Figure 44.4. The PRED vs. DV plot shows a tendency toward overprediction at the lower end of the concentration scale. It is even more obvious on the IPRED vs. DV plot. Overall, the trough levels are slightly overpredicted for the nelfinavir treatment group. However, no obvious trend is observed in the distribution of WRES vs. PRED and/or time.
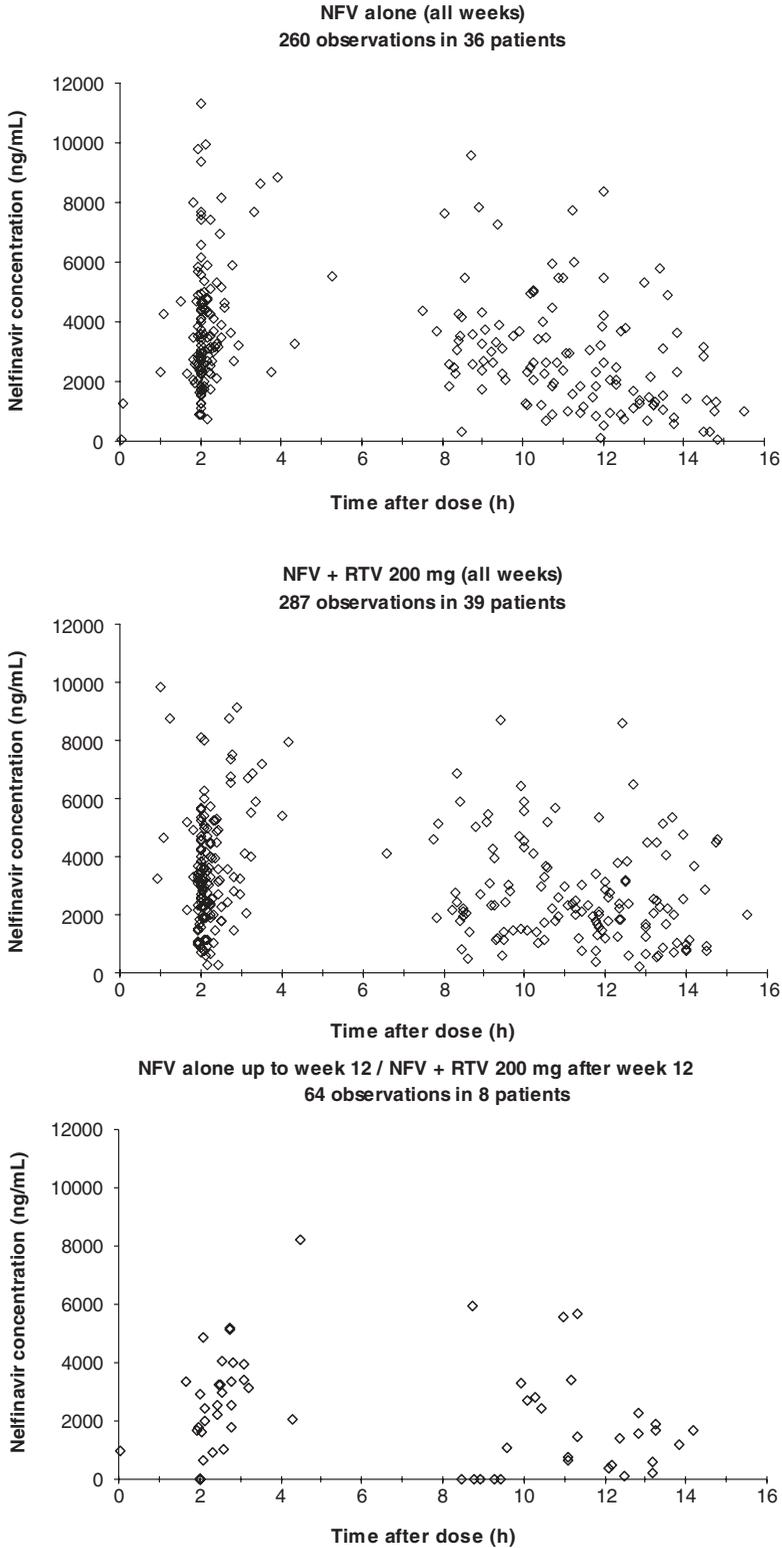
FIGURE 44.2 Nelfinavir concentration–time data by treatment group.
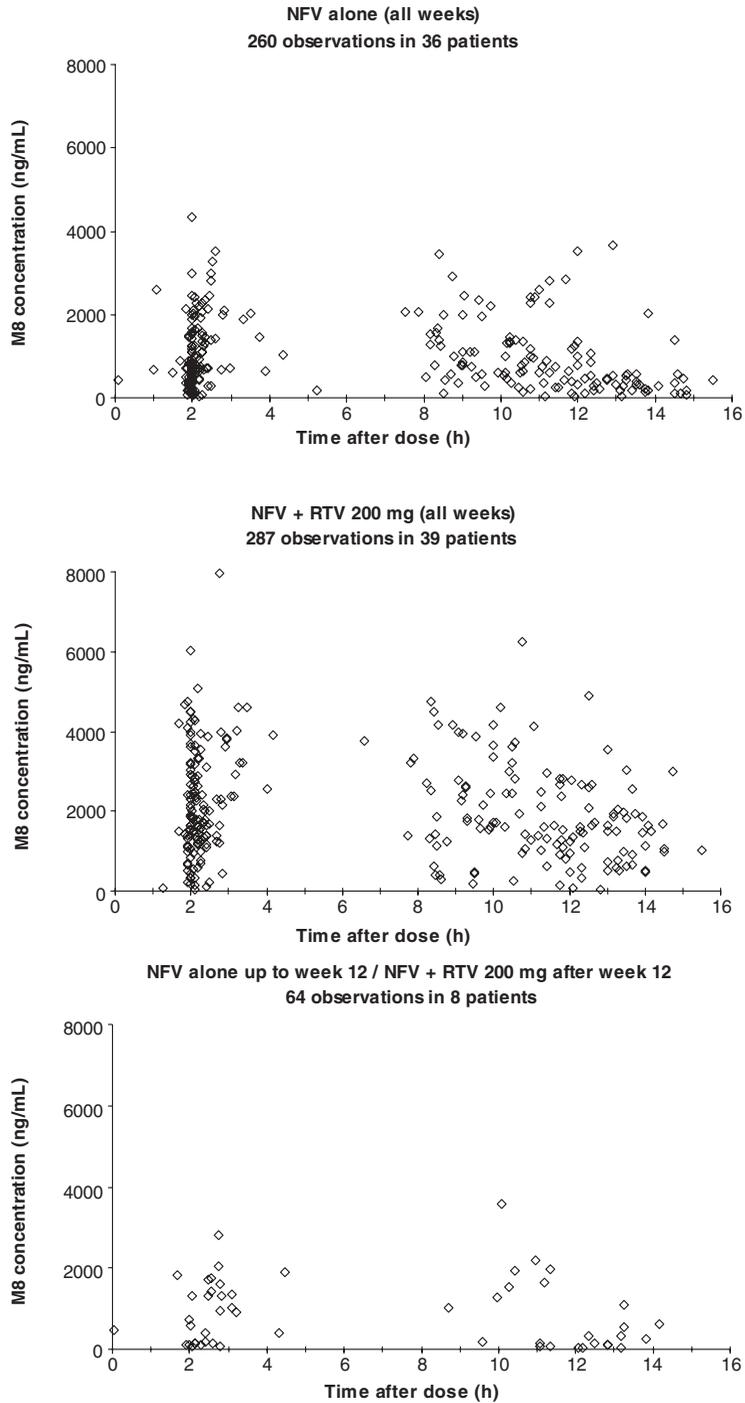
FIGURE 44.3 M8 concentration–time data by treatment group.

TABLE 44.2   Population Pharmacokinetics of Nelfinavir Fitted Alone

Parameter	Population Estimate	Relative Standard Error (%)
Clearance, CL (L/h)	31.9	5.36
Volume of distribution, V (L)	72.4	18.2
Absorption rate constant, KA (h^{-1})	0.0994	16.4
Lag time, ALAG1 (h)	0.921	1.17
Between-subject variance of CL (%)	40.9	23.2
Between-subject variance of V (%)	66.9	36.9
Between-occasion variance of CL (%)	38.6	21.2
Proportional residual error (%)	25.6	13.7
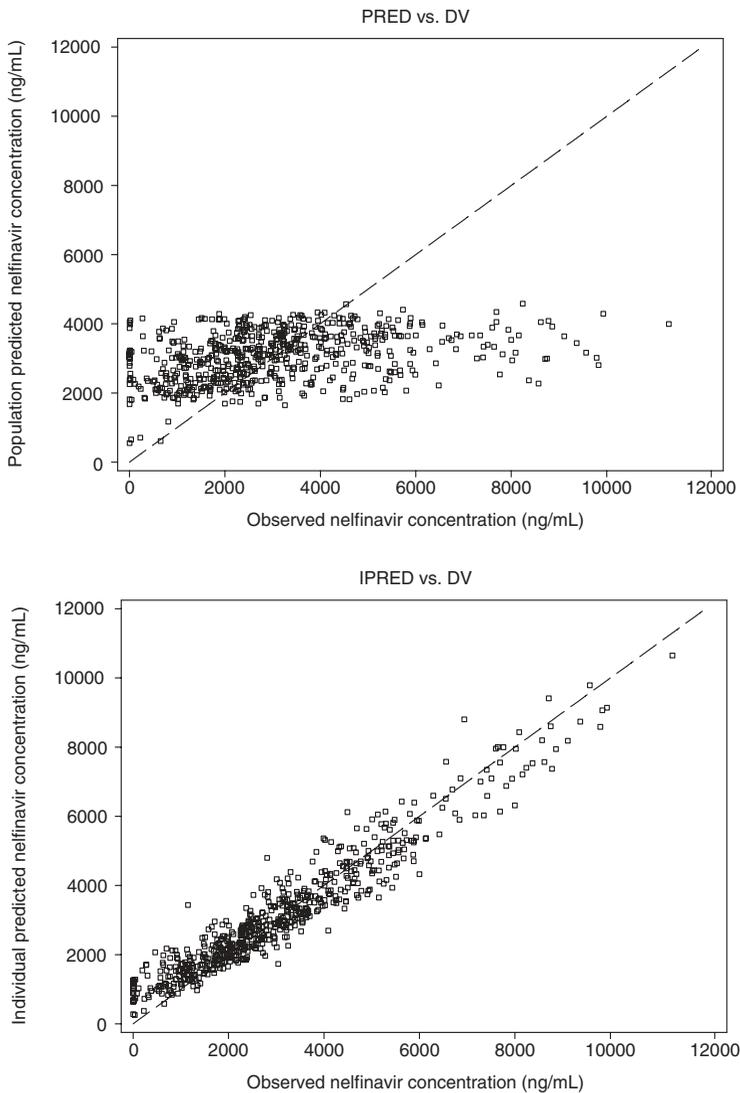
FIGURE 44.4   Goodness-of-fit plots of population PK model for nelfinavir alone.
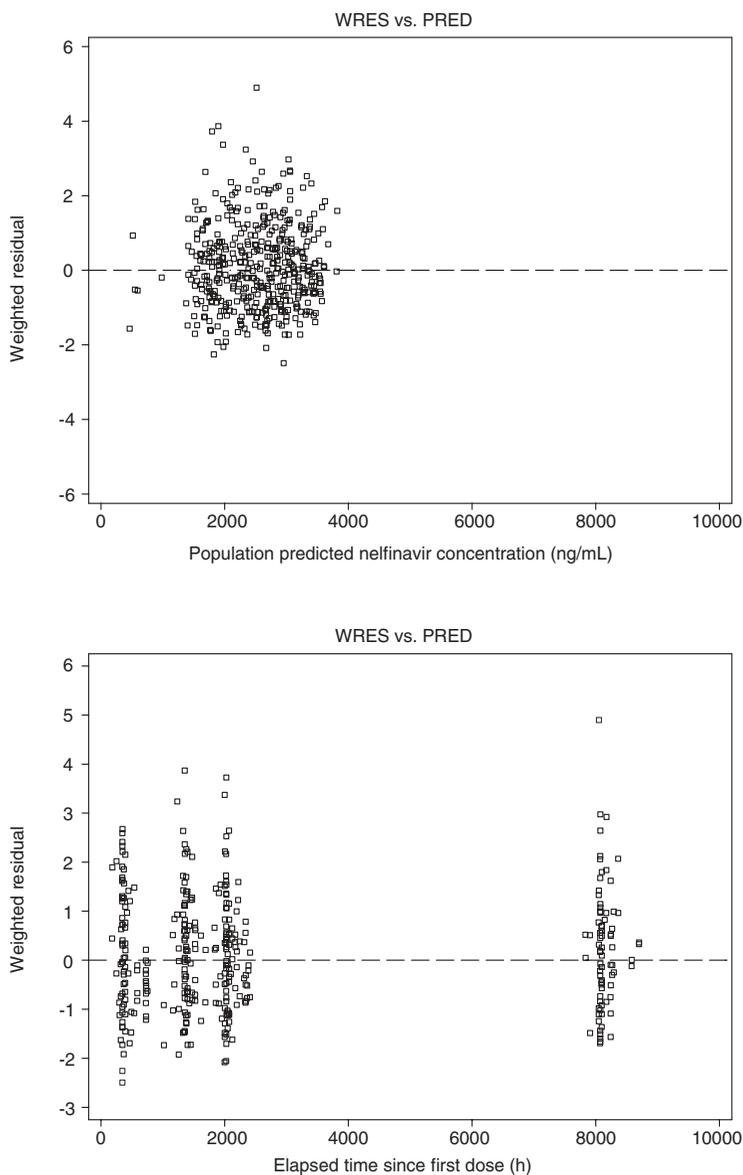
FIGURE 44.4   *Continued*

44.2.2.3   *Nelfinavir and M8*

Both between-subject and interoccasion variances were estimated on clearance of nelfinavir, absorption rate constant, and clearance of M8. The residual error with a proportional error model was modeled for nelfinavir and M8 separately. The effect of ritonavir was found to have a statistically significant impact on the clearance of M8 but not on that of nelfinavir. The apparent clearance of M8 was 3.23 L/h; it decreased to 1.87 L/h when nelfinavir was coadministered with ritonavir. After univariate selection, a large number of covariates were included in the full model. According to the acceptance criteria, none of the effect on clearance of nelfinavir on

clearance of M8 remained in the final model after backward deletion. The population parameters of nelfinavir and M8 (coding 1) are presented in Table 44.3.

After fitting the nelfinavir and M8 data simultaneously, the half-life was no longer driven by the absorption rate constant. The terminal half-file was equal to 7.6 h. Simultaneous modeling of nelfinavir and M8 provided a population PK profile for nelfinavir similar to the one obtained when the nelfinavir data were fitted alone.

The goodness-of-fit plots for nelfinavir and M8, separately, are presented in Figure 44.5. The tendency toward overprediction at the lower end of the concentration

TABLE 44.3   Population Pharmacokinetics of Nelfinavir and M8 (Coding 1)

Parameter	Population Estimate	Relative Standard Error (%)
Clearance of nelfinavir, CL (L/h)	31.5	5.27
Volume of distribution of nelfinavir, V (L)	345	16.5
Absorption rate constant of nelfinavir, KA (h^{-1})	0.481	14.4
Lag time of nelfinavir, ALAG1 (h)	0.921	0.492
Rate constant of formation of M8, K23 (h^{-1})	0.00287	29.5
Clearance of M8, CLM (L/h)	3.23	26.3
Ritonavir coadministration effect on CLM (L/h)	−1.36	34.4
Between-subject variance of CL (%)	37.5	23.8
Between-subject variance of KA (%)	72.0	36.5
Between-subject variance of CLM (%)	60.2	30.3
Between-occasion variance of CL (%)	36.6	16.5
Between-occasion variance of KA (%)	81.3	31.6
Between-occasion variance of CLM (%)	44.5	28.4
Proportional residual error for NFV (%)	16.8	20.9
Proportional residual error for M8 (%)	29.4	22.2
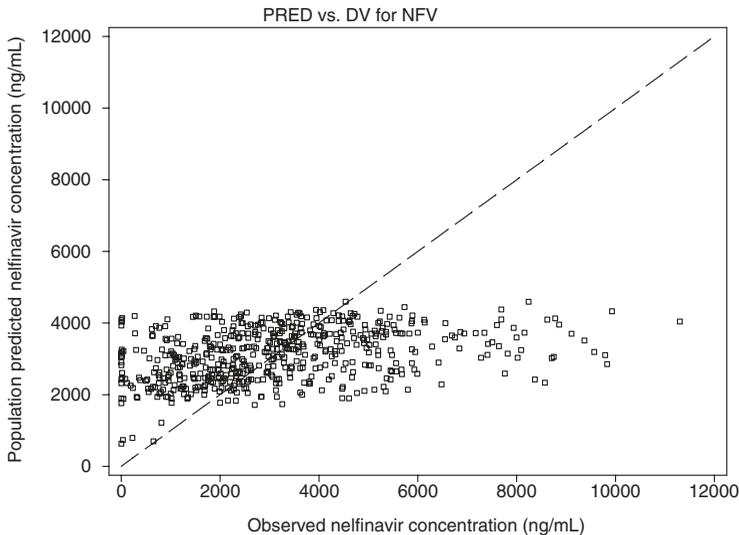
FIGURE 44.5   Goodness-of-fit plots of population PK model for nelfinavir and M8 nelfinavir data.
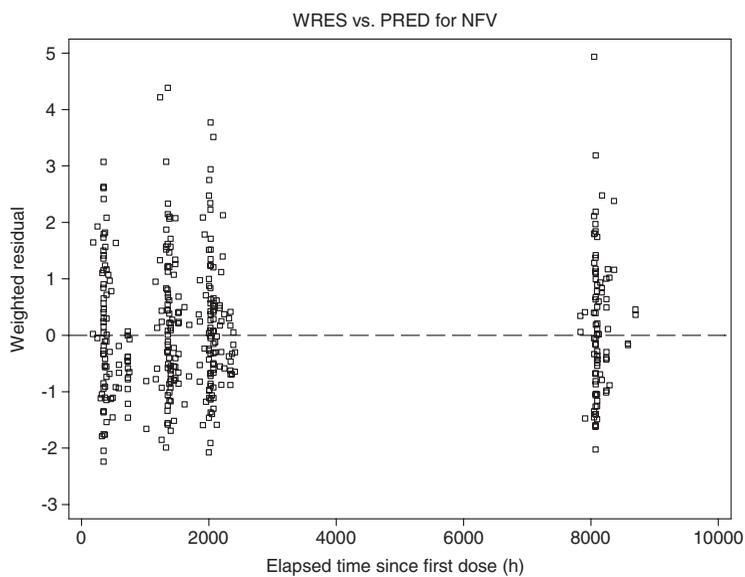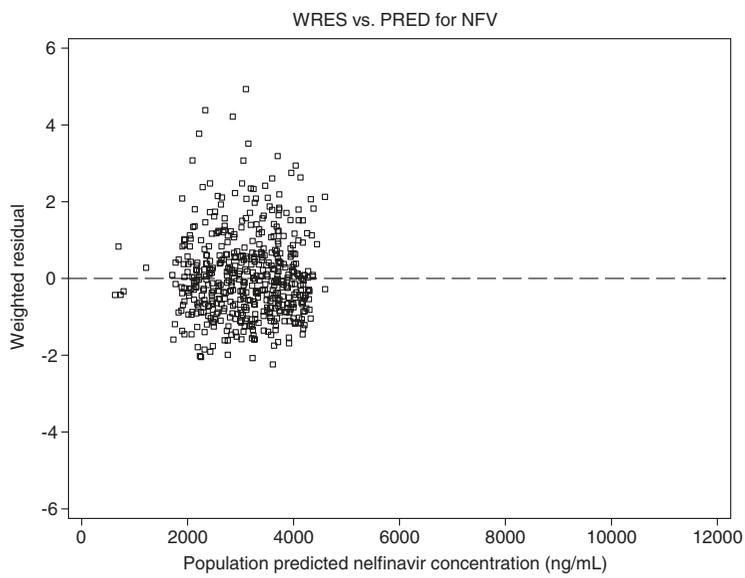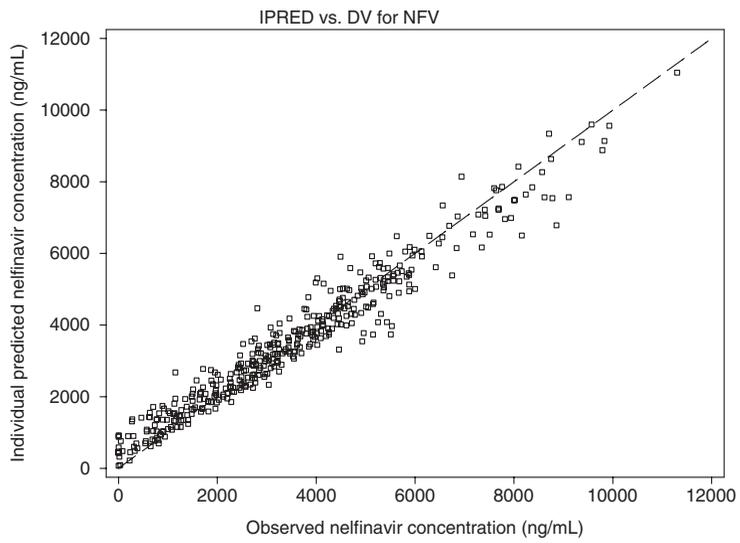
FIGURE 44.5 *Continued*
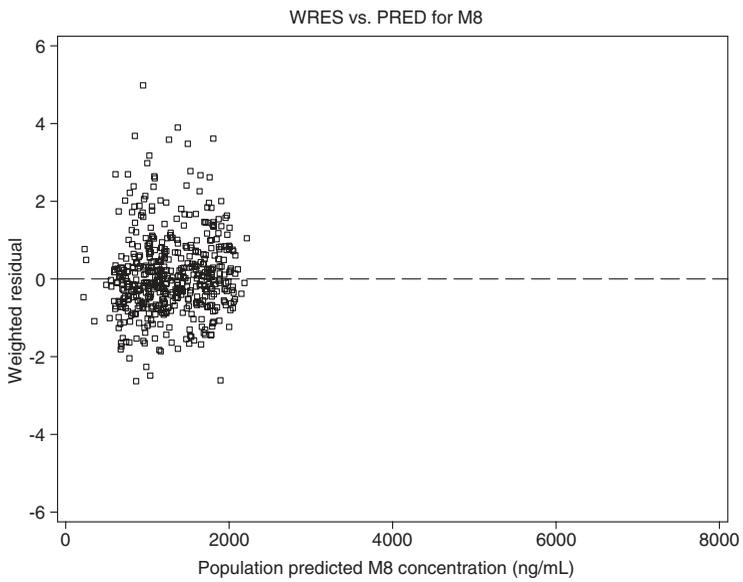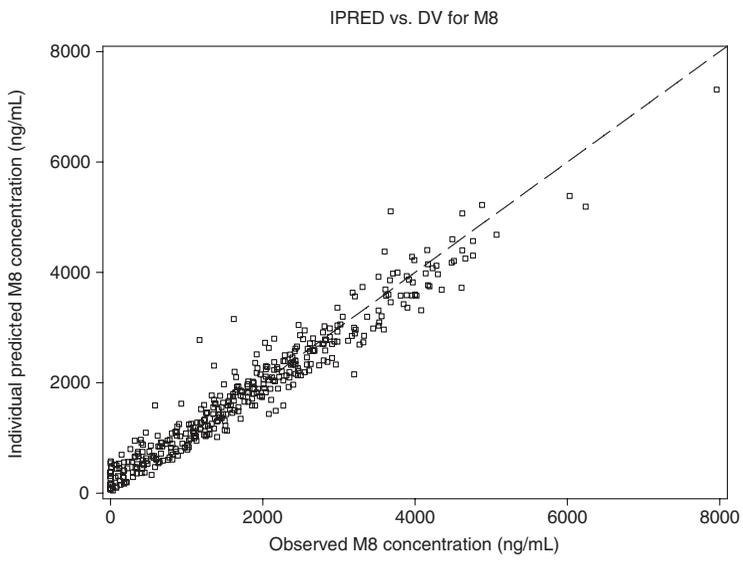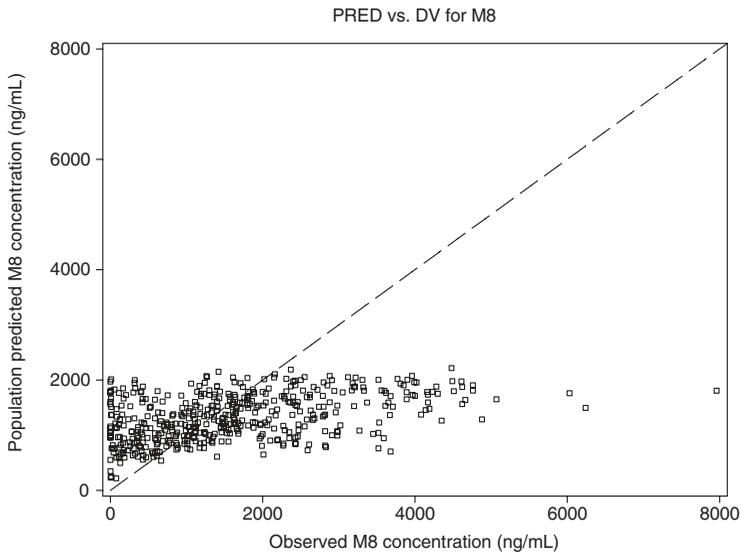
FIGURE 44.5 *Continued*
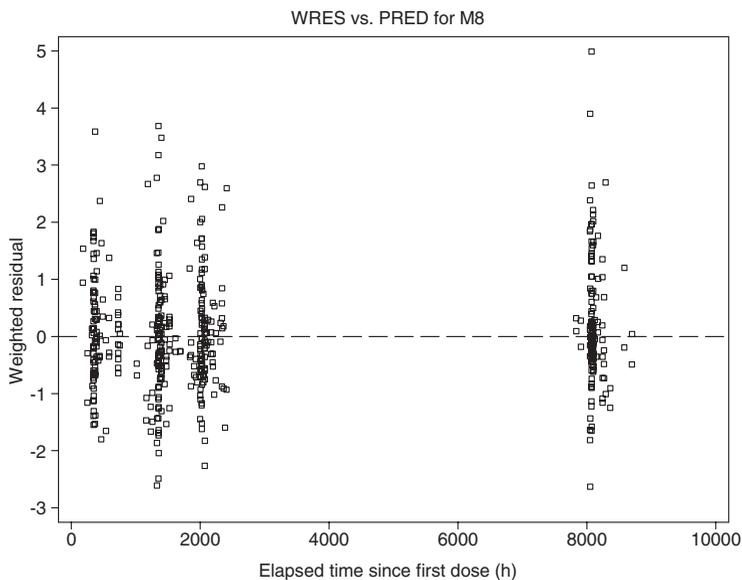
FIGURE 44.5 *Continued*

TABLE 44.4 Geometric Mean of Exposure Parameters of Nelfinavir for the Nelfinavir and the Nelfinavir Plus Ritonavir Treatment Groups

Treatment Weeks	AUC_τ of NFV (h·ng/mL)		C_{12} of NFV (ng/mL)		C_{max} of NFV (ng/mL)	
	NFV	NFV+RTV	NFV	NFV+RTV	NFV	NFV+RTV
2	35,868	35,917	1,948	2,049	3,920	3,808
8	33,635	35,791	1,809	2,056	3,676	3,724
12	37,865	33,952	2,177	1,849	3,984	3,699
48	41,917	38,606	2,280	2,200	4,569	4,140

TABLE 44.5 Geometric Mean of Exposure Parameters of M8 for the Nelfinavir and the Nelfinavir Plus Ritonavir Treatment Groups

Treatment Weeks	AUC_τ of M8 (h·ng/mL)		C_{12} of M8 (ng/mL)		C_{max} of M8 (ng/mL)	
	NFV	NFV+RTV	NFV	NFV+RTV	NFV	NFV+RTV
2	9,541	22,250	551	1,416	1,066	2,369
8	8,788	20,214	503	1,289	982	2,125
12	9,887	20,440	605	1,247	1,065	2,226
48	12,483	24,289	725	1,548	1,389	2,599

scale, seen on IPRED vs. DV plots for nelfinavir alone, is less marked when nelfinavir and M8 are modeled together.

The geometric mean of the derived exposure parameters are presented in Table 44.4 for nelfinavir and in Table 44.5 for M8. Exposure parameters of nelfinavir were

not different between the nelfinavir alone and the nelfinavir plus ritonavir treatment groups. Exposure parameters of M8 were more than doubled in the nelfinavir plus ritonavir treatment group compared to the nelfinavir treatment group.

44.2.2.4 Comparison Between Models
The population parameters obtained with the different coding 1–4 are presented in Table 44.6. Regardless of the coding used, the results were consistent with similar values of clearance, volume, and absorption parameters for nelfinavir.

When fixing the distribution volume of M8 to 1 L, the different coding 1, 3, and 4 gave similar values for the clearance of M8. The microrate constant K23 is related to FMET by Eq (44.3). The FMET calculated with the coding 1 is equal to 0.0314, which is in concordance with value estimated with the coding 3.

When fixing the distribution volume of M8 equal to the volume of nelfinavir, the microrate constant K23 was 17% lower than the one with coding 1. This decrease was compensated by a 15% decrease of the clearance of M8, which resulted in the same AUC of M8 for the two coding approaches. The percentage of decrease of the clearance of M8 with ritonavir was identical for the two coding approaches, being 42%. The relative standard errors of the estimates were larger with the coding 2, above 50% for the M8 parameters.

44.2.3 Discussion

A one-compartment model with first-order absorption and elimination with a lag time adequately described the nelfinavir PK profile. Although the nelfinavir elimination is mainly metabolic, a model with a Michaelis–Menten elimination was not considered because (a) only one dose of nelfinavir was tested leading to a concentration range between peak and trough not large enough to estimate Michaelis–Menten parameters; and (b) there were no signs of saturation of the elimination on the plots. No saturable mechanism has been reported in the literature for nelfinavir.

Two samples were collected at different weeks of treatment: the predose samples were collected between 7 and 16 hours after the evening dose and the second samples, planned to be collected between 2 and 6 hours after the morning dose, were all collected 2 hours post morning dose. Since the C_{max} of nelfinavir at steady state is achieved 4–5 hours after administration, the information necessary to accurately estimate the volume and the absorption rate constant was missing in patients. Therefore, volume and absorption rate constant could not be easily identified separately and different sets of parameter values lead to the same values of clearance, apparent half-life, and prediction. The absorption of nelfinavir was slow compared to the elimination, with comparable half-lives (33). Consequently, in the absence of intravenous data, flip-flop phenomena may have been present during modeling.

Nelfinavir is known to have different trough levels after morning or evening administration. Those variations have been observed either in healthy subjects or patients. An increase of 46% in clearance of nelfinavir between morning and evening administration was estimated for healthy subjects. This diurnal effect seems to disappear when ritonavir is coadministered (data on file). The sampling scheme in that study prevents the evaluation of that effect in patients. The impossibility to evaluate the diurnal effect on clearance in patients, when nelfinavir was administered alone, could lead to an underestimation of the clearance and therefore an

TABLE 44.6 Population Pharmacokinetics of Nelfinavir and M8: Comparison Between Coding

Parameter	Estimates Coding 1	Estimates Coding 2	Estimates Coding 3	Estimates Coding 4
Clearance of nelfinavir, CL (L/h)	31.5	32.0	31.7	31.6
Volume of distribution of nelfinavir, V (L)	345	348	336	347
Absorption rate constant of nelfinavir, KA (h^{-1})	0.481	0.445	0.568	0.482
Lag time of nelfinavir, $ALAG1$ (h)	0.921	0.744	1.16	0.921
Rate constant of formation of M8, $K23$ (h^{-1}) or fraction of nelfinavir clearance for M8 formation, $FMET^{a}$	0.00287	0.833	0.0357	0.00286
Clearance of M8, CLM (L/h)	3.23	953	3.29	3.25
Ritonavir coadministration effect on CLM (L/h)	−1.36	−399	−1.46	−1.37
Between-subject variance of CL (%)	37.5	37.9	38.1	37.5
Between-subject variance of KA (%)	72.0	73.7	68.6	72.4
Between-subject variance of CLM (%)	60.2	61.2	46.3	60.4
Between-occasion variance of CL (%)	36.6	37.5	37.1	36.7
Between-occasion variance of KA (%)	81.3	85.1	76.9	81.7
Between-occasion variance of CLM (%)	44.5	45.2	50.4	44.6
Proportional residual error for NFV (%)	16.8	15.7	17.6	16.8
Proportional residual error for M8 (%)	29.4	29.9	26.5	29.5

^{a} FMET was calculated with coding 3.

overprediction of the concentrations. The plots of the population prediction and/or the individual predictions versus the observations showed a tendency toward overprediction at the lower end of the concentration scale, which was more obvious in the nelfinavir alone treatment group. This tendency was also accentuated by concentrations in patients below 360 ng/mL, that is, the lowest concentrations recorded in healthy subjects, which were not compatible with the pharmacokinetics of nelfinavir under steady-state conditions unless the compliance had been poor or the time and dosing history wrong. Those concentrations, 36 in total, represent 5.9% of the patient data; among those, 18 BLQ concentrations were replaced with half of quantification limit.

The population clearance of nelfinavir estimated with nelfinavir and M8 data modeled together was identical to that with nelfinavir data alone, 31.5 versus 31.9 L/h. The distribution volume of nelfinavir was different, 345 versus 72.4 L, compensated by a different value of absorption rate constant, 0.481 versus $0.0994\,h^{-1}$. The flip-flop phenomenon already mentioned when modeling nelfinavir data alone was no longer present when M8 data were added. In fact, the elimination half-life calculated for the nelfinavir and M8 data modeled together was comparable to the apparent half-life, that is, absorption half-life, for nelfinavir data alone, 7.6 and 7.0 h, respectively. The two sets of parameters for nelfinavir resulted in similar nelfinavir concentration–time profiles. The absence of flip-flop phenomenon with the full model for nelfinavir and M8 shows that metabolite data can help the modeling of the parent compound by at least increasing the number of degrees of freedom. For the same reason, the residual variance for nelfinavir decreased from 25.6% to 16.8% when adding M8 data. The advantage of the full model over separated modeling of parent drug and metabolite is advocated even under various assumptions (34). The comparison between the different coding shows that the estimated parameters were similar.

Overall, the pharmacokinetics of nelfinavir was less variable than the pharmacokinetics of M8 with a between-subject variance for clearance of 37.5% for nelfinavir compared to 60.2% for M8 and a proportional residual error of 16.8% for nelfinavir compared to 29.4% for M8.

Ritonavir is a potent CYP 3A4 inhibitor and is used at low doses to elevate plasma concentrations of other protease inhibitors being primarily metabolized by CYP 3A4. This type of interaction has proved advantageous for saquinavir since saquinavir exposure and thus its efficacy was increased (4). The combination with nelfinavir showed much smaller effects on nelfinavir levels; it appears to change the M8/nelfinavir concentration ratio from 0.3 to 0.6 in normal metabolizers (5). While CYP 3A4 is responsible for more than 90% of the hepatic metabolism of saquinavir, it is not the only one involved in the elimination of nelfinavir. The decrease of CYP 2C19 activity in chronic liver disease is accompanied with low nelfinavir clearance and decreased M8 formation, since CYP 2C19 catalyzes roughly 50% of nelfinavir clearance (35). The analysis of patient data shows that the ritonavir coadministration does not modify the nelfinavir clearance, likely through a compensation mechanism with the other cytochrome P450 enzymes, but decreases the M8 clearance by 42%.

Exposure parameters (i.e., AUC_τ, C_{max}, C_{12h}) calculated for nelfinavir and M8 reflected the effect of ritonavir coadministration on the metabolism of nelfinavir. The M8 exposure parameters were increased by 87–157% when nelfinavir

was coadministered with ritonavir, while nelfinavir exposure parameters remained unchanged. The M8/nelfinavir AUC_τ ratio (without molecular weight correction) changed from 0.26–0.30 when nelfinavir was administered alone to 0.56–0.63 when nelfinavir was administered with ritonavir. The agreement of those results with literature information (29) confirms the ability of the model to estimate the clearance of nelfinavir and M8 and thus the ability to derive individual AUC_τ.

44.3 SUMMARY

The nelfinavir example presents the simultaneous modeling of parent drug and metabolite in a large data set. Since the metabolite has not been administered alone, the modeling of the parent drug and its metabolite requires simplification so that metabolite parameters could be estimated. This consists of using the ratio of the rate of conversion to the metabolite volume and apparent elimination rate constant for the metabolite. Simultaneous modeling of nelfinavir and M8 allows for a better description of the data through a more stable model. In addition, it was possible to evaluate the impact of the coadministration of ritonavir.

REFERENCES

1. D. E. Drayer, Pharmacologically active drug metabolites: therapeutic and toxic activities, plasma and urine data in man, accumulation in renal failure, in *Handbook of Clinical Pharmacokinetics* M. Gibaldi and L. Prescott (Eds.). ADIS Health Science Press, Sydney, 1983, pp. 114–132.

2. R. Xie, R. H. Mathijssen, A. Sparreboom, J. Verweij, and M. O. Karlsson, Clinical pharmacokinetics of irinitecan and its metabolites in relation with diarrhea. *Clin Pharmacol Ther* **72**:265–275 (2002).

3. J. V. Gobburu, V. Tammara, L. Lesko, S. S. Jhee, J. J. Sramek, N. R. Cutler, and R. Yuan, Pharmacokinetic–pharmacodynamic modeling of rivastigmine, a cholinesterase inhibitor, in patients with Alzheimer's disease. *J Clin Pharmacol* **41**:1082–1090 (2001).

4. P. N. Zannikos, S. Rohatagi, B. K. Jensen, S. L. DePhillips, and G. R. Rhodes, Pharmacokinetics and concentration–effect analysis of intravenous RGD891, a platelet GPIIb/IIIa antagonist, using mixed-effects modeling (NONMEM). *J Clin Pharmacol* **40**:1129–1140 (2000).

5. I. Ragueneau, C. Laveille, R. Jochemsen, G. Resplandy, C. Funck-Bretano, and P. Jaillon, Pharmacokinetic–pharmacodynamic modeling of the effects of ivabradine, a direct sinus node inhibitor, on heart rate in healthy volunteers. *Clin Pharmacol Ther* **64**:192–203 (1998).

6. H. J. Zimmerman and W. C. Maddrey, Acetaminophen (paracetamol) hepatotoxicity with regular intake of alcohol: analysis of instances of therapeutic misadventure. *Hepatology* **22**:767–773 (1995).

7. F. V. Shiodt, F. A. Rochling, D. L. Casey, and W. M. Lee, Acetaminophen toxicity in an urban county hospital. *N Engl J Med* **337**:1112–1117 (1997).

8. *Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling.* CDER, Nov. 1999.

9. T. Kerbusch, R. L. H. Jansen, R. A. A. Mathot, A. D. R. Huitma, M. Jansen, R. E. N. van Rijswijk, and J. H. Beijnen, Modulation of the cytochrome P450-mediated metab-

olism of ifosfamide by ketoconazole and rifampin. *Clin Pharmacol Ther* **70**:132–141 (2001).

10. W. M. Sallas, S. Milosavljev, J. D'souza, and M. Hossain, Pharmacokinetic drug interactions in children taking oxcarbazepine. *Clin Pharmacol Ther* **74**:138–149 (2003).

11. D. Callies, D. P. de Alwis, J. G. Wright, A. Sandler, M. Burgess, and L. Aarons, A population pharmacokinetic model for doxorubicin and doxorubicinol in the presence of a novel MDR modulator, zosuquidar trihydrochloride (LY335979). *Cancer Chemother Pharmacol* **51**:107–118 (2003).

12. J. C. Panetta, M. N. Kirstein, A. Gajjar, G. Nair, M. Fouladi, R. L. Heideman, M. Wilkinson, and C. F. Stewart, Population pharmacokinetics of temozolomide and metabolites in infants and children with primary central nervous system tumors. *Cancer Chemother Pharmacol* **52**:435–441 (2003).

13. T. Kerbusch, J. de Kraker, R. A. Mathjt, and J. H. Beijne, Population pharmacokinetics of ifosfamide and its dechloroethylated and hydroxilated metabolites in children with malignant disease: a sparse sampling approach. *Clin Pharmacokinet* **40**:615–625 (2001).

14. J. Lotsch, C. Skarke, H. Schmidt, J. Liefhold, and G. Geisslinger, Pharmacokinetic modeling to predict morphine and morphine-6-glucuronide plasma concentrations in healthy young volunteers. *Clin Pharmacol Ther* **72**:151–162 (2002).

15. C. E. Klein, E. Gupta, J. M. Reid, P. J. Atherton, J. A. Sloan, H. C. Pitot, M. J. Ratain, and H. Kastrissios, Population pharmacokinetic model for irinotecan and two of its metabolites, SN-38 an SN-38 glucurinide. *Clin Pharmacol Ther* **72**:638–647 (2002).

16. T. Kerbusch, R. A. Mathot, H. J. Keizer, G. P. Kaijser, J. H. Schellens, and J. H. Beijnen, Influence of dose and infusion duration on pharmacokinetics of ifosfamide and metabolites. *Drug Metab Dispos* **29**:967–975 (2001).

17. T. Kerbusch, U. Walby, P. A. Milligan, and M. O. Karlsson, Population pharmacokinetic modelling of darifenacin and its hydroxylated metabolite using pooled data, incorporating saturable first-pass metabolism, CYP2D6 genotype and formulation-dependent bioavailability. *Br J Clin Pharmacol* **56**:639–652 (2003).

18. C. H. Koks, A. D. Huitema, E. D. Kroon, T. Chuenyam, R. W. Sparidans, J. M. Lange, and J. H. Beijnen, Population pharmacokinetics of itraconazole in Thai HIV-1-infected persons. *Ther Drug Monit* **25**:229–233 (2003).

19. N. E. Schoemaker, R. A. Mathot, P. Schoffski, H. Rosing, J. H. Schellens, and J. H. Beijnen, Development of an optimal pharmacokinetic sampling schedule for rubitecan administered orally in a daily times five schedule. *Cancer Chemother Pharmacol* **50**:514–517 (2002).

20. A. D. R. Huitema, R. A. A. Mathot, M. M. Tibben, J. H. M. Schellens, S. Rodenhuis, and J. H. Beijnen, Population pharmacokinetics of thioTEPA and its active metabolite TEPA in patients undergoing high-dose chemotherapy. *Br J Clin Pharmacol* **51**:61–70 (2001).

21. M. Levi, D. A. Dempsey, N. L. Benowitz, and L. B. Sheiner, Population pharmacokinetic model for nicotine (NIC) metabolism to cotinine (COT) and 3-hydroxy cotinine (3HC) in healthy volunteers. ASCPT Annual Meeting, Washington DC, April 2–5, 2003. Abstract, *Clin Pharmacol Ther* **73**:21 (2003).

22. M. Hassan, U. S. Svensson, P. Ljungman, B. Bjorkstrand, H. Olsson, M. Bielentein, M. Abdel-Rehim, C. Nilsson, M Johansson, and M. O. Karlsson, A mechanism-based pharmacokinetic-enzyme model for cyclophosphamide autoinduction in breast cancer patients. *Br J Clin Pharmacol* **48**:669–677 (1999).

23. K. E. Zhang, E. Wu, A. K. Patik, B. Kerr, M. Zorbas, A. Lankford, T. Kobayashi, Y. Maeda, B. Shetty, and S. Webber, Circulating metabolites of the human immunodefi-

ciency virus protease inhibitor nelfinavir in humans: structural identification, levels in plasma, and antiviral activities. *Antimicrob Agents Chemother* **45**:1086–1093 (2001).

24. M. H. Zhang, Y. K. Pithavala, C. A. Lee, J. H. Lillibridge, E. Y. Wu, T. M. Sandoval, R. G. Daniels, and B. M. Kerr, Apparent genetic polymorphism in nelfinavir metabolism: evaluation of clinical relevance. *Proceedings of the 12th International Symposium of Microsomes and Drug Oxidations*, Montpellier, France, July 20–24, 1998, abstract 264.

25. Viracept®, NDA 20-778/SLR 020, NDA 20-779/SLR 041, NDA 21-503/SLR 002, Agouron Pharmacauticals, Inc., revised May 2, 2003.

26. A. Bardsley-Elliot and G. L. Plosker, Nelfinavir: an update on its use in HIV infection. *Drugs* **59**:581–620 (2000).

27. G. Skowron, G. Leoung, B. Kerr, A. Dusek, R. Anderson, S. Beebe, and R. Grosso, Lack of pharmacokinetic interaction between nelfinavir and nevirapine. *AIDS* **12**:1243–1244 (1998).

28. M. Kurowski, M. Müller, F. Donath, M. Mrozikiewicz, and C. Möcklinghoff, Single daily doses of saquinavir achieve HIV-inhibitory concentrations when combined with "BABY-DOSE" ritonavir. *Eur J Med Res* **4**:101–104 (1999).

29. P. A. Baede-van Dijk, P. W. Hugen, C. P. Verwij-van Wissen, P. P. Koopmans, D. M. Burger, and Y. A. Hekster, Analysis of variation in plasma concentrations of nelfinavir and its active metabolite M8 in HIV-positive patients. *AIDS* **15**:991–998 (2001).

30. J. G. Gerber, Using pharmacokinetics to optimize antiretroviral drug–drug interactions in the treatment of human immunodeficiency virus infection. *Clin Infect Dis* **30**:S123–129 (2000).

31. K. A. Jackson, S. E. Rosenbaum, B. M. Kerr, Y. K. Pithavala, G. Yuen, and M. N. Dudley, A population pharmacokinetic analysis of nelfinavir mesylate in human immunodeficiency virus-infected patients enrolled in a phase III clinical trial. *Antimicrob Agents Chemother* **44**:1832–1837 (2000).

32. A. J. Boeckman, L. B. Sheiner, and S. L. Beal, *NONMEM Users Guides.* NONMEM Project Group, San Francisco, 1992 and 1998.

33. M. Pfister, L. Labbé, S. M. Hammer, J. Mellors, K. K. Bennett, and S. Rosenkranz, Population pharmacokinetics and pharmacodynamics of efavirenz, nelfinavir, and indinavir: adults AIDS clinical trial group study 398. *Antimicrob Agents Chemother* **47**:130–137 (2003).

34. *NONMEM UsersNet Archives: Metabolite Modeling.* `htpp://www.cognigencorp.com/nonmen/nm` (December 1999).

35. Y. Khaliq, K. Gallicano, I. Seguin, K. Fyke, G. Carignan, D. Bulman, A. Badley, and D. W. Cameron, Single and multiple dose pharmacokinetics of nelfinavir and CYP 2C19 activity in human immunodeficiency virus-infected patients with chronic liver disease. *Br J Clin Pharmacol* **50**:108–115 (2000).

APPENDIX 44.1   CODING 1

```
$PROB POP PK NFV & M8 PATIENTS

$INPUT STN ID TRT WEEK DAT1=DROP TIME AMPM AMT DV EVID MDV SS II
CM.T TYP

$DATA PK_NFVandM8_PAT.CSV IGNORE S
$SUBROUTINE ADVAN5
```

```
$MODEL
COMP=(DEPOT, DEFDOSE)
COMP=(CENTRAL, DEFOBS)
COMP=(METABOL)

$PK
RTV=1
IF (TRT.EQ.2) RTV=0
IF (TRT.EQ.3.AND.WEEK.LE.12) RTV=0

OC1=0
IF (WEEK.LE.2) OC1=1
OC2=0
IF (WEEK.EQ.8) OC2=1
OC3=0
IF (WEEK.EQ.12) OC3=1
OC4=0
IF (WEEK.EQ.48) OC4=1

;NELFINAVIR PARAMETERS
TVCL=THETA(1)
CL=TVCL*EXP(ETA(1)+OC1*ETA(4)+OC2*ETA(5)+OC3*ETA(6)+OC4*ETA(7))

V2=THETA(2)

KA=THETA(3)*EXP(ETA(2)+OC1*ETA(8)+OC2*ETA(9)+OC3*ETA(10)+OC4*ETA
(11))
ALAG1=THETA(4)

;M8 PARAMETERS
V3=1
K23=THETA(5)
CLM1=THETA(6)+RTV*THETA(7)
CLM=CLM1*EXP(ETA(3)+OC1*ETA(12)+OC2*ETA(13)+OC3*ETA(14)+OC4*ETA
(15))

K12=KA
K20=(CL/V2)-K23
K30=CLM/V3

S2=V2/1000
S3=V3/1000

$ERROR
DEL=0
IF (F.EQ.0) DEL=0.0001

;PROPORTIONAL ERROR
```

```
Y1=(F+DEL)*(1+ERR(1))  ;NFV
Y2=(F+DEL)*(1+ERR(2))  ;M8


Y=TYP*Y1+(1-TYP)*Y2


IPRED=F


$THETA
  (0,30)      ;1 APP CLEARANCE OF NFV
  (0,100)     ;2 APP CENTRAL VOL
  (0,0.3)     ;3 1ST ORD ABS
  (0,1)       ;4 LAGTIME
  (0,0.05)    ;5 K23
  (0,3)       ;6 APP CLEARANCE OF M8
  (-5,-1,5)   ;7 RTV ON CLM


$SIGMA
0.05 0.05              ;PROPORTIONAL ERROR
$OMEGA
0.01 0.01 0.01
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME


$ESTIMATION MAX=9999 METHOD=0 PRINT=5 NOABORT POSTHOC
$COVARIANCE
```

APPENDIX 44.2   CODING 2

```
$PROB POP PK NFV & M8 PATIENTS - ALTERNATIVE WRITING

$INPUT STN ID TRT WEEK DAT1=DROP TIME AMPM AMT DV EVID MDV SS II
CMT TYP

$DATA PK_NFVandM8_PAT.CSV IGNORE S

$SUBROUTINE ADVAN5
```

```
$MODEL
COMP=(DEPOT, DEFDOSE)
COMP=(CENTRAL, DEFOBS)
COMP=(METABOL)

$PK
RTV=1
IF (TRT.EQ.2) RTV=0
IF (TRT.EQ.3.AND.WEEK.LE.12) RTV=0

OC1=0
IF (WEEK.LE.2) OC1=1
OC2=0
IF (WEEK.EQ.8) OC2=1
OC3=0
IF (WEEK.EQ.12) OC3=1
OC4=0
IF (WEEK.EQ.48) OC4=1

;NELFINAVIR PARAMETERS
TVCL=THETA(1)
CL=TVCL*EXP(ETA(1)+OC1*ETA(4)+OC2*ETA(5)+OC3*ETA(6)+OC4*ETA(7))

V2=THETA(2)

KA=THETA(3)*EXP(ETA(2)+OC1*ETA(8)+OC2*ETA(9)+OC3*ETA(10)+OC4*ETA
(11))
ALAG1=THETA(4)

;M8 PARAMETERS
V3=V2
K23=THETA(5)
CLM1=THETA(6)+RTV*THETA(7)
CLM=CLM1*EXP(ETA(3)+OC1*ETA(12)+OC2*ETA(13)+OC3*ETA(14)+OC4*ETA
(15))

K12=KA
K20=(CL/V2)-K23
K30=CLM/V3

S2=V2/1000
S3=V3/1000

$ERROR
DEL=0
IF (F.EQ.0) DEL=0.0001

;PROPORTIONAL ERROR
```

```
Y1=(F+DEL)*(1+ERR(1))  ;NFV
Y2=(F+DEL)*(1+ERR(2))  ;M8


Y=TYP*Y1+(1-TYP)*Y2


IPRED=F


$THETA
  (0,30)         ;1 APP CLEARANCE OF NFV
  (0,200)        ;2 APP CENTRAL VOL
  (0,0.3)        ;3 1ST ORD ABS
  (0,1)          ;4 LAGTIME
  (0,0.5)        ;5 K23
  (0,100)        ;6 APP CLEARANCE OF M8
  (-500,-50,500) ;7 RTV ON CLM
$SIGMA
0.05 0.05          ;PROPORTIONAL ERROR
$OMEGA
0.01 0.01 0.01
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME


$ESTIMATION MAX=9999 METHOD=0 PRINT=5 NOABORT POSTHOC
$COVARIANCE
```

APPENDIX 44.3   CODING 3

```
$PROB POP PK NFV & M8 PATIENTS - ALTERNATIVE WRITING

$INPUT STN ID TRT WEEK DAT1=DROP TIME AMPM AMT DV EVID MDV SS II
CMT TYP

$DATA PK_NFVandM8_PAT.CSV IGNORE S

$SUBROUTINE ADVAN5
```

```
$MODEL
COMP=(DEPOT, DEFDOSE)
COMP=(CENTRAL, DEFOBS)
COMP=(METABOL)

$PK
RTV=1
IF (TRT.EQ.2) RTV=0
IF (TRT.EQ.3.AND.WEEK.LE.12) RTV=0

OC1=0
IF (WEEK.LE.2) OC1=1
OC2=0
IF (WEEK.EQ.8) OC2=1
OC3=0
IF (WEEK.EQ.12) OC3=1
OC4=0
IF (WEEK.EQ.48) OC4=1

;NELFINAVIR PARAMETERS
FMET=THETA(5)

TVCL=THETA(1)
CL=TVCL*EXP(ETA(1)+OC1*ETA(4)+OC2*ETA(5)+OC3*ETA(6)+OC4*ETA(7))

V2=THETA(2)

KA=THETA(3)*EXP(ETA(2)+OC1*ETA(8)+OC2*ETA(9)+OC3*ETA(10)+OC4*ETA
(11))
ALAG1=THETA(4)

;M8 PARAMETERS
V3=1

CLM1=THETA(6)+RTV*THETA(7)
CLM=CLM1*EXP(ETA(3)+OC1*ETA(12)+OC2*ETA(13)+OC3*ETA(14)+OC4*ETA(15))

K12=KA
K20=(1-FMET)*CL/V2
K23=FMET*CL/V2
K30=CLM/V3

S2=V2/1000
S3=V3/1000

$ERROR
DEL=0
IF (F.EQ.0) DEL=0.0001
```

```
;PROPORTIONAL ERROR

Y1=(F+DEL)*(1+ERR(1))  ;NFV
Y2=(F+DEL)*(1+ERR(2))  ;M8

Y=TYP*Y1+(1-TYP)*Y2

IPRED=F

$THETA
  (0,30)       ;1 APP CLEARANCE OF NFV
  (0,100)      ;2 APP CENTRAL VOL
  (0,0.3)      ;3 1ST ORD ABS
  (0,1)        ;4 LAGTIME
  (0,0.05,1)   ;5 FMET
  (0,3)        ;6 APP CLEARANCE OF M8
  (-5,-1,5)    ;7 RTV ON CLM

$SIGMA
0.05 0.05            ;PROPORTIONAL ERROR
$OMEGA
0.01 0.01 0.01
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME

$ESTIMATION MAX=9999 METHOD=0 PRINT=5 NOABORT POSTHOC
$COVARIANCE
```

APPENDIX 44.4   CODING 4

```
$PROB POP PK NFV & M8 PATIENTS - ALTERNATIVE WRITING

$INPUT STN ID TRT WEEK DAT1=DROP TIME AMPM AMT DV EVID MDV SS II
CMT TYP

$DATA PK_NFVandM8_PAT.CSV IGNORE S

$SUBROUTINE ADVAN6 TRANS=1 TOL=4
```

```
$MODEL
COMP=(DEPOT, DEFDOSE)
COMP=(CENTRAL, DEFOBS)
COMP=(METABOL)

$PK
RTV=1
IF (TRT.EQ.2) RTV=0
IF (TRT.EQ.3.AND.WEEK.LE.12) RTV=0

OC1=0
IF (WEEK.LE.2) OC1=1
OC2=0
IF (WEEK.EQ.8) OC2=1
OC3=0
IF (WEEK.EQ.12) OC3=1
OC4=0
IF (WEEK.EQ.48) OC4=1

;NELFINAVIR PARAMETERS

CL=THETA(1)*EXP(ETA(1)+OC1*ETA(4)+OC2*ETA(5)+OC3*ETA(6)+OC4*ETA
(7))

V2=THETA(2)

KA=THETA(3)*EXP(ETA(2)+OC1*ETA(8)+OC2*ETA(9)+OC3*ETA(10)+OC4*ETA
(11))
ALAG1=THETA(4)

;M8 PARAMETERS
V3=1

K23=THETA(5)
CLM1=THETA(6)+RTV*THETA(7)
CLM=CLM1*EXP(ETA(3)+OC1*ETA(12)+OC2*ETA(13)+OC3*ETA(14)+OC4*ETA(15))

K12=KA
K20=(CL/V2)-K23
K30=CLM/V3

S2=V2/1000
S3=V3/1000

;CALCULATION OF TIME AFTER DOSE
;IF (EVID.EQ.1) THEN
;TDOS=TIME
;TAD=0
;ENDIF
;IF (EVID.NE.1) TAD=TIME-TDOS
```

```
$DES

DADT(1)=-KA*A(1)
DADT(2)=KA*A(1)-K20*A(2)-K23*A(2)
DADT(3)=K23*A(2)-K30*A(3)

$ERROR
DEL=0
IF (F.EQ.0) DEL=0.0001

;PROPORTIONAL ERROR
Y1=(F+DEL)*(1+ERR(1))  ;NFV
Y2=(F+DEL)*(1+ERR(2))  ;M8

Y=TYP*Y1+(1-TYP)*Y2

;IPRED=F

$THETA
  (0,30)      ;1 APP CLEARANCE OF NFV
  (0,100)     ;2 APP CENTRAL VOL
  (0,0.3)     ;3 1ST ORD ABS
  (0,1)       ;4 LAGTIME
  (0,0.05)    ;5 K23
  (0,3)       ;6 APP CLEARANCE OF M8
  (-5,-1,5)   ;7 RTV ON CLM

$SIGMA
0.05 0.05              ;PROPORTIONAL ERROR
$OMEGA
0.01 0.01 0.01
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) 0.05
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME

$ESTIMATION MAX=9999 METHOD=0 PRINT=5 NOABORT
$COVARIANCE
```

Characterizing Nonlinear Pharmacokinetics: An Example Scenario for a Therapeutic Protein

STUART FRIEDRICH

45.1 INTRODUCTION

The pharmacokinetics of a therapeutic agent may be considered nonlinear based on several characteristics that may be observed in exposure data. In the simplest case, the relationship between the observed exposure, whether it be maximum concentration (C_{max}), area under the concentration–time curve (AUC), or some other exposure metric, does not change linearly with respect to the administered dose. In other cases, the observed exposure changes over time during repeat administration of the therapeutic agent. Saturation of processes that govern the absorption, metabolism, distribution, or elimination of a therapeutic agent is a common cause of observing nonlinear pharmacokinetics. It is also possible for a therapeutic agent or coadministered agents to induce or suppress processes that subsequently change the observed pharmacokinetics.

When nonlinear pharmacokinetics are observed, it is important to characterize the nonlinearity to determine if it could adversely affect safety or efficacy and to choose appropriate dosing regimens for future studies and labeling. Unexpected future observations that delay or terminate development of the therapeutic agent may result if nonlinearity is inadequately characterized or incorrectly deemed to be caused by a certain factor. Sources (1, 2) and some modeling approaches (3–6) for nonlinearity observed with small molecules have been covered elsewhere. For protein therapeutics there are two common sources of nonlinearity: (a) the development of antibodies against the protein (7) and (b) binding of the protein to the target receptor (8–11). In addition to these two sources, monoclonal antibodies (mAbs) may also occasionally show nonlinearity with respect to dose if a very broad dose range is studied that includes doses high enough to saturate Fc receptor-dependent processes that are responsible for distribution and elimination (12). In

the example discussed below, a mAb is observed to have dose-dependent pharma-cokinetics, but also produces an antibody response in some patients and exhibits nonlinear pharmacokinetics in antibody-positive patients. A population analysis of the data is performed using two types of models and these models are compared in terms of their ability to describe the data, as well as the advantages and disad-vantages of each model.

45.2   AN EXAMPLE

45.2.1   Study Design

The data used in this example are not actual observed data, but rather were simu-lated using a hierarchical model and a set of parameters that generated data that was representative for a typical mAb. The simulation model was structurally different from both of the models that were used to analyze the data to minimize any bias when comparing the ability of each model to fit the data. The mAb was assumed to generate a nonneutralizing anti-mAb response in some subjects and was also assumed to bind to target receptors expressed on circulating cells (i.e., antigen presenting cells or effector cells).

Data were analyzed from two simulated studies, representing typical Phase 1 and Phase 2 studies. In the Phase 1 study, 40 subjects were randomized to receive a single subcutaneous dose of 0.03, 0.1, 0.3, 1, or 3 mg/kg of mAb (8 subjects per group). Concentrations of mAb were obtained at 0, 1, 2, 4, 8, 12, and 24 hours and 2, 3, 5, 7, 14, 21, 28, 35, and 42 days postdose. Anti-mAb antibody titers were obtained 1, 2, 4, 6, and 8 weeks postdose. The Phase 2 study included 200 patients randomized to receive 0.05, 0.2, 0.75, or 2 mg/kg of mAb subcutaneously (50 patients per group). Each patient received one dose every 4 weeks for a total of 24 weeks (6 doses). Concentrations of mAb were obtained 3, 14, and 28 days after each dose. Anti-mAb antibody titers were obtained 4 weeks after each dose. The lower quan-titation limit for mAb concentrations and anti-mAb titers was 0.1 ng/mL and 1.4 (log titer), respectively, for both studies. The assay for mAb was performed using plasma; therefore, this assay only quantified mAb that was unbound to the target receptor present on circulating cells.

45.2.2   Model Development

To guide model development, the observed data were first examined graphically to determine general characteristics and to look for trends with respect to dose, time, and the impact of anti-mAb antibodies. Models were developed using NONMEM (Version 5). Two different model types were developed: the first model (MODEL 1, see Appendix 45.1) used an analytical solution (closed-form) where the nonlinear-ity was accounted for by allowing the model parameters to be a function of mAb dose and the titer of anti-mAb antibody, while the second model (MODEL 2, see Appendix 45.2) used differential equations to allow a more mechanistic approach to characterize the nonlinearity. For each model, three estimation methods were evaluated: first-order (FO), first-order conditional estimation (FOCE), and FOCE with interaction. Various forms of between-subject variability models were evalu-

ated in each of the tested models, including additive versus proportional and uni-variate versus multivariate variability models. Body weight, dose, and anti-mAb titer were the only covariates that were examined in the models. Selection of the most appropriate model was based on agreement between predicted and observed plasma concentrations, randomness in the weighted residuals versus the predicted values, convergence of the estimation and covariance routines, reasonable para-meter and error estimates, good precision of the parameter and error estimates, and decreases in the objective function ($-2 \cdot$log likelihood of the data; -2LL) of ≥ 6.635 points ($p < 0.01$) when comparing nested models within MODEL 1 or MODEL 2.

45.2.3 Anti-mAb and Target Receptor Binding Effects

The incidence of anti-mAb-positive patients and the titer of the response are shown in Table 45.1. The incidence and titer did not appear to vary with dose. Anti-mAb was first detected at week 2 in the Phase 1 study. There was a wide variation in the time course of the anti-mAb response in the Phase 2 study, where the titers in most patients peaked and then declined slowly over the study, while in other subjects the titers either continued to increase or peaked and remained constant.

Patients who had detectable anti-mAb titers had a rate of mAb elimination that appeared to increase as mAb concentrations decreased and was also dependent on the anti-mAb titer (Figure 45.1).

The wide variation in the kinetics of the anti-mAb response and lack of sufficient anti-mAb data prevented the development of a model that could be used to predict the anti-mAb response in anti-mAb-positive patients. It was also possible that the presence of mAb caused interference in the anti-mAb assay. For these reasons, the mAb model used a simplified metric of anti-mAb levels to account for the impact of anti-mAb on mAb elimination. In both studies, the anti-mAb titer measured 4 weeks after dosing was found to be a reasonable covariate that predicted both the

TABLE 45.1 Incidence and Titer of Anti-mAb-Positive Patients in Phase 1 and 2 Studies

Phase 1 Study			Phase 2 Study			
Dose (mg/kg)	Incidence (n (%))	Log$_{10}$ Titer at Week 4 (individual values)	Dose (mg/kg)	Incidence (n (%))	Log$_{10}$ Titer at Week 4 (mean (range))	
					Dose 1	Dose 6[a]
0.03	0 (0)		0.05	10 (20)	1.78 (1.40–2.71)	0.00 (0.00–0.00)
0.1	2 (25)	1.51, 1.75	0.2	13 (26)	2.01 (1.40–2.78)	0.00 (0.00–0.00)
0.3	4 (50)	2.06, 1.88, 1.61, 1.95	0.75	16 (32)	1.90 (1.40–2.39)	1.35 (0.00–2.16)
1.0	2 (25)	1.85, 2.54	2.0	11 (22)	1.92 (1.40–2.49)	1.16 (0.00–2.07)
3.0	1 (12.5)	1.69				

[a]Summary statistics include those subjects who had anti-mAb titers that fell below detection by Dose 6.
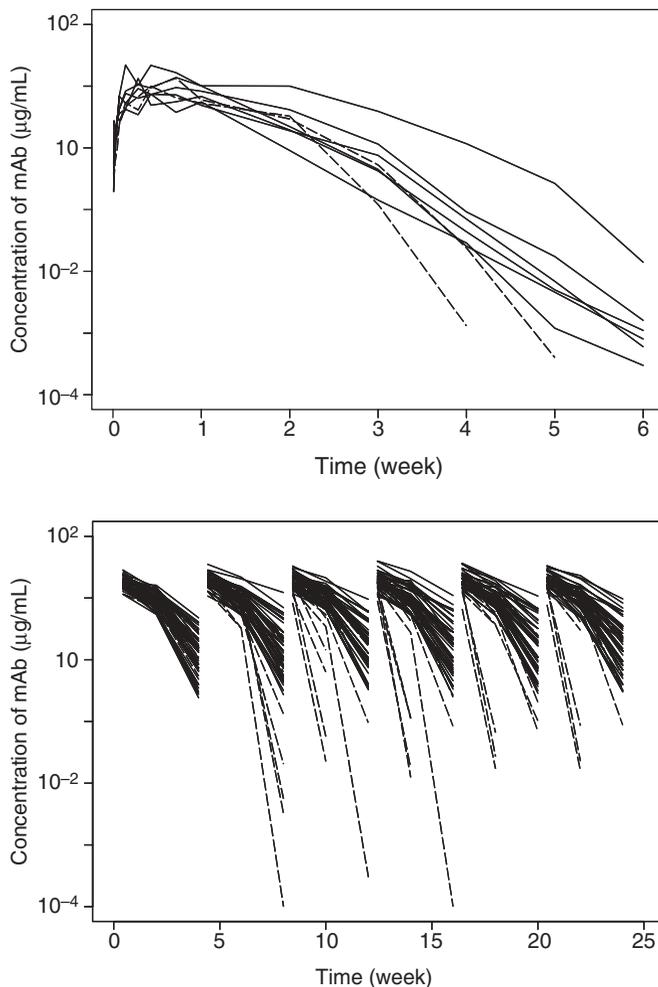
FIGURE 45.1 Concentration of mAb versus time for subjects in Phase 1 study who received 3 mg/kg dose (upper panel) and for patients in Phase 2 study who received 2 mg/kg dose (lower panel). Subjects with negative or positive anti-mAb response shown with solid or dashed lines, respectively.

durability and strength of the anti-mAb response, and also reduced the possibility of mAb interference in the anti-mAb assay. In the Phase 2 study, the anti-mAb titer covariate for each dosing interval was set equal to the anti-mAb titer observed at the end of the dosing interval, thus accounting for changes in the anti-mAb titer during multiple dosing. Although the anti-mAb titers were reported as the log of the titer, the models used the actual titer as the covariate value. To account for the delay in anti-mAb response, the mAb models also assumed a constant lag time of 1 week after the first dose before anti-mAb had any impact on mAb elimination.

Independent of the effect of anti-mAb on mAb elimination, the maximum concentration and elimination rate were also found to have a nonlinear relationship with dose. This was observed in both the Phase 1 and Phase 2 studies for both anti-
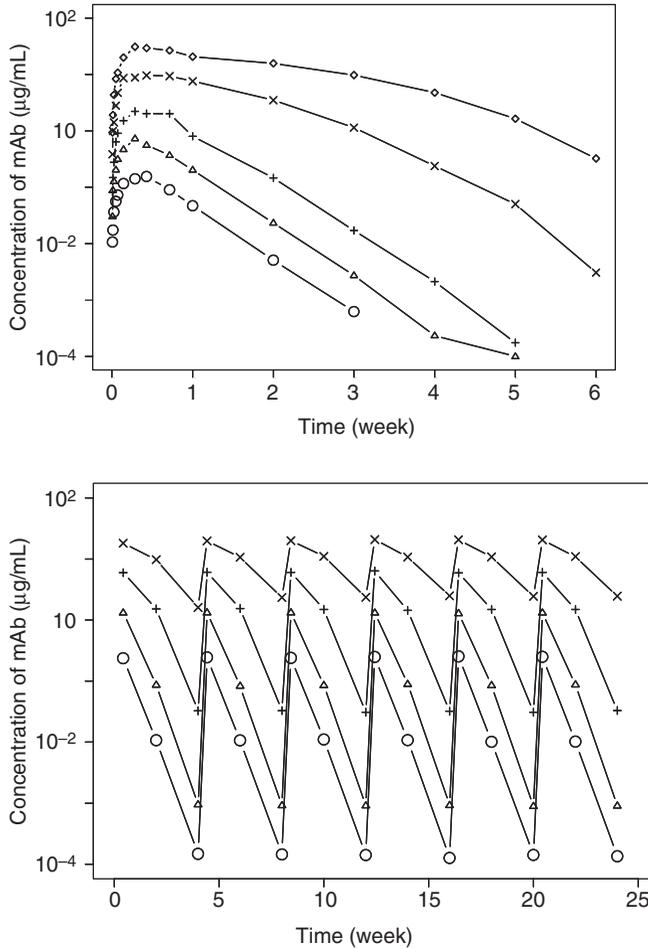
FIGURE 45.2 Mean concentration of mAb versus time for subjects in Phase 1 (upper panel) and Phase 2 (lower panel). Dose levels were 0.03(○), 0.1(Δ), 0.3(+), 1(×), and 3 mg/kg (◊) in the Phase 1 study and 0.05(○), 0.2(Δ), 0.75(+), and 2 mg/kg (×) in the Phase 2 study. These plots do not include anti-mAb-positive subjects, indicating nonlinear pharmacokinetics were present even in anti-mAb-negative subjects.

mAb-positive and anti-mAb-negative subjects and appeared to be independent of time and number of doses (Figure 45.2).

45.2.4   Final Developed Models

Based on the criteria specified under Model Development (Section 45.2.2), the best basic model structure found for both MODEL 1 and MODEL 2 was one compartment with first-order absorption. Observed data were modeled as follows:

$$C_{ij} = f(\theta_i, D_i, t_{ij}) \cdot \varepsilon_{ij} \tag{45.1}$$

where f is the structural model, C_{ij} is the jth observed concentration in the ith individual, θ_i is the set of model parameters for the ith individual, D_i is the dose received by the ith individual, t_{ij} is the time from dose for the jth observation, and ε_{ij} is the residual error between the observed and model predicted concentration, assumed to be normally distributed with a mean of zero and variance of σ^2.

MODEL 1 was implemented in NONMEM using ADVAN 2 (TRANS2) and the following model parameter equations:

$$V = \theta_1 \cdot BWT \cdot Dose^{\theta_5} \cdot e^{\eta_1} \tag{45.2}$$

$$CL = \theta_2 \cdot Dose^{\theta_6} \cdot e^{\eta_2} \quad \text{for } LATR = 0 \tag{45.3}$$

$$CL = \theta_2 \cdot Dose^{\theta_6} \cdot e^{\eta_2} + \theta_4 \cdot 10^{LATR} \cdot e^{\eta_4} \quad \text{for } LATR > 0 \tag{45.4}$$

$$K_a = \theta_3 \cdot e^{\eta_3} \tag{45.5}$$

where V is effective volume of distribution (mL), CL is clearance (mL/h), K_a is the oral absorption rate constant (h^{-1}), BWT is body weight (kg), and $LATR$ is the $\log_{10}$ of the anti-mAb titer. As shown, power functions for V and CL incorporating dose were found to best represent the nonlinear pharmacokinetics in MODEL 1, while the impact of anti-mAb was incorporated as an additive linear increase in CL relative to anti-mAb titers. Proportional intersubject variability was included on V, CL, and K_a. V was found to increase in proportion to body weight. The scalar used to translate model predicted amounts to concentration was set equal to V. A proportional residual error model was found to best represent the data. MODEL 1 was fit to a data set that contained the data from both the Phase 1 and Phase 2 studies. No study impact was found on any model parameters.

The basic structure of MODEL 2 is shown in Figure 45.3. MODEL 2 was implemented using ADVAN6 (TRANS1) and the following model differential equations:

$$\frac{\partial A_1}{\partial t} = -K_a \cdot A_1 \tag{45.6}$$
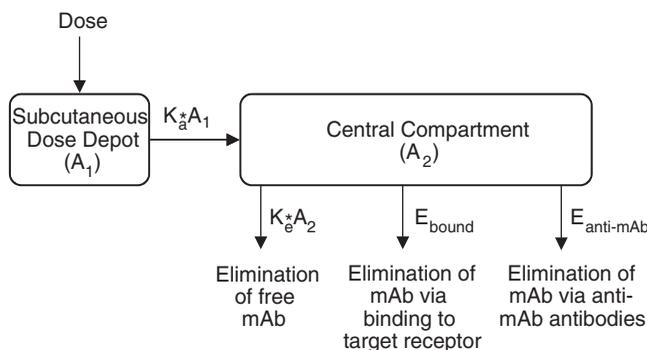
FIGURE 45.3 MODEL 2 diagram. MODEL 2 was mechanistic in nature and included three elimination routes for mAb, where both E_{bound} and $E_{anti-mAb}$ had a nonlinear relationship with concentration of mAb.

$$\frac{\partial A_2}{\partial t} = K_a \cdot A_1 \cdot f_u - K_e \cdot A_2 - E_{\text{anti-mAb}} - E_{\text{bound}} \tag{45.7}$$

where

$$E_{\text{anti-mAb}} = \frac{V_{\text{max}} \cdot C_2}{C_2 + K_m} \tag{45.8}$$

$$E_{\text{bound}} = K_{e1} \cdot C_{\text{bound}} \cdot V \tag{45.9}$$

$$f_u = \frac{A_2}{A_2 + C_{\text{bound}} \cdot V} \tag{45.10}$$

$$C_{\text{bound}} = \frac{C_2 \cdot C_{\text{target},t=0}}{K_D + C_2} \tag{45.11}$$

$$C_2 = \frac{A_2}{V} \tag{45.12}$$

In the above equations A_1 was the amount of mAb in the subcutaneous dose compartment (μmol), A_2 was the amount of mAb not bound to target receptor or free mAb in the central compartment (μmol), C_2 was the concentration of free mAb in the central compartment (μmol/mL), K_a was the subcutaneous absorption rate constant (h^{-1}), K_e was the elimination rate constant for free mAb (h^{-1}), f_u was the free fraction of mAb (not bound to the target receptor), and K_D was the equilibrium dissociation constant for binding between mAb and target receptors (μmol/mL). Included in the model are also terms that represent the elimination due to anti-mAb ($E_{\text{anti-mAb}}$) and elimination due to binding to the target receptor (E_{bound}). Elimination via anti-mAb was found to be best represented using the Michaelis–Menten equation (Eq. (45.8)) to capture the dependence of elimination on the concentration of mAb. The elimination via binding to target receptors (Eq. (45.9)) was modeled using a first-order rate constant K_{e1} (h^{-1}), where elimination was proportional to the concentration of mAb bound to target receptors (C_{bound}). Preliminary structures for MODEL 2 (see Appendix 45.2) included differential equations to calculate the concentration of bound and unbound mAb using forward and reverse binding rate constants for the mAb; however, these model equation systems were found to be stiff and computationally resource intensive. In the final model above, an assumption was made to allow a simpler set of model equations. For most mAb, the forward and reverse binding rates to the target receptor are much greater than the rates of absorption or elimination of the mAb, and this was also true for the mAb in this case. So the assumption applied in MODEL 2 was that unbound mAb was in equilibrium with bound mAb at all times. This assumption eliminated the need to include separate equations for bound and free target receptors, and bound mAb in the differential mass balance equations. However, since the elimination of mAb via binding to target receptors was proportional to C_{bound}, the model required C_{bound} expressed in terms of C_2. This is provided by Eq. (45.11), which is derived by combining the equilibrium binding equation (Eq. (45.13)) with an equation relating the total concentration of target sites to unbound target sites (Eq. (45.14)), where $C_{\text{target},t=0}$ is the concentration of unbound target sites at time zero.

$$K_D = \frac{C_2 \cdot C_{\text{target free}}}{C_{\text{bound}}} \tag{45.13}$$

$$C_{\text{target,bound}} = C_{\text{target},t=0} - C_{\text{bound}} \tag{45.14}$$

Equations (45.13) and (45.14) assume that the mAb binds stoichiometrically with the target receptor in a 1:1 ratio, and Eq. (45.14) assumes that the total number of target receptors remains constant. Since the binding to target receptors was assumed to be in equilibrium at all times, binding was also taken into account when mAb is transferred from the dose compartment to the central compartment, with the central compartment only accounting for free mAb that is in equilibrium with bound mAb. Although not immediately apparent, these model structures preserve the overall mass balance in the model.

MODEL 2 also used the following parameter equations:

$$F_1 = 1/150,000 \tag{45.15}$$

$$K_a = \theta_1 \cdot e^{\eta_1} \tag{45.16}$$

$$V_{\max} = 0 \quad \text{for } LATR = 0 \tag{45.17}$$

$$V_{\max} = \frac{\theta_2}{150,000} \cdot 10^{LATR} \quad \text{for } LATR > 0 \tag{45.18}$$

$$K_m = \frac{\theta_3}{150,000} \tag{45.19}$$

$$K_e = \theta_4 \cdot e^{\eta_2} \tag{45.20}$$

$$K_{e1} = \theta_5 \tag{45.21}$$

$$V = \theta_6 \cdot BWT \cdot e^{\eta_3} \tag{45.22}$$

where F_1 was fixed but used to convert the dose amount from μg to μmol, using the mAb molecular weight of 150,000 daltons. The maximum rate ($V_{\max}$, μmol/h) and Michaelis–Menten constant (K_m, μmol/mL) for elimination via anti-mAb were divided by 150,000 so that the model θ values were expressed in terms of more relevant mass units rather than molar units. Similar to MODEL 1, V was found to increase in proportion to body weight. The scalar used to translate model predicted amounts in μmol to concentrations in μg/mL was set equal to V/150,000. Proportional intersubject variability was included on V, K_e, and K_a. A proportional residual error model was found to best represent the data. MODEL 2 was fit first to the Phase 1 data alone and then to both the Phase 1 and Phase 2 data combined. The population value and intersubject variance for K_a determined from fitting the Phase 1 data were fixed during fitting of the combined Phase 1 and 2 data. K_D and $C_{\text{target},t=0}$ could not be estimated and were fixed to 1×10^{-5} (μmol/mL) and 2×10^{-6} (μmol/mL), respectively, which were expected to represent typical values for a mAb and cell surface target receptor.

45.2.5   Model Comparison and Parameter Values

Tables 45.2 and 45.3 list the parameter values for each model, while Figures 45.4 and 45.5 show some diagnostic plots for each model. Overall, both models were

TABLE 45.2   Final Model Parameters for MODEL 1

Parameter	Parameter Description	Estimate (Value ± SE)
θ_1	Volume of distribution per kg of body weight (mL/kg)	50.0 ± 1.58
θ_2	Coefficient for dose-dependent clearance term (mL/h)	29.6 ± 1.05
θ_3	Subcutaneous absorption rate (h^{-1})	0.0132 ± 0.000161
θ_4	Coefficient for anti-mAb-dependent clearance (mL/h)	0.0244 ± 0.00512
θ_5	Volume of distribution linearity with respect to dose (power model)	0.156 ± 0.0395
θ_6	Clearance linearity with respect to dose (power model)	−0.457 ± 0.0243
ω_1	Intersubject variance of volume of distribution	0.0688 ± 0.0159
ω_2	Intersubject variance of dose-dependent clearance term	0.0977 ± 0.0145
ω_3	Intersubject variance of subcutaneous absorption	0.00568 ± 0.00230
ε	Residual error variance	0.0853 ± 0.0103
	Objective function value	−12413

TABLE 45.3   Final Model Parameters for MODEL 2

Parameter	Parameter Description	Estimate (Value ± SE)
θ_1	Subcutaneous absorption rate (h^{-1})	0.0387*
θ_2	Maximum elimination rate coefficient for anti-mAb-dependent elimination ($\mu g/h$)	0.418 ± 0.0382
θ_3	Michaelis–Menten constant for anti-mAb-dependent elimination ($\mu g/mL$)	4.87 ± 0.363
θ_4	Elimination rate constant for free mAb (h^{-1})	0.00132 ± 0.000155
θ_5	Elimination rate constant for mAb bound to target receptors (h^{-1})	0.0600 ± 0.000831
θ_6	Volume of distribution per kg of body weight (mL/kg)	88.9 ± 6.53
ω_1	Intersubject variance of subcutaneous absorption	0.0468[a]
ω_2	Intersubject variance of elimination rate constant for free mAb	0.172 ± 0.0855
ω_3	Intersubject variance of volume of distribution	0.0256 ± 0.00415
ε	Residual error variance	0.0595 ± 0.00827
	Objective function value	−13624

[a] Subcutaneous absorption parameters were estimated using the Phase 1 data and subsequently fixed when fitting the combined Phase 1 and Phase 2 data.

able to fit the data reasonably well, with MODEL 2 having slightly lower residual error values compared to MODEL 1. When the data were fit using MODEL 1 and MODEL 2 without accounting for anti-mAb effects (θ_4 set to zero in MODEL 1, θ_2 set to zero in MODEL 2), the objective function values increased by approximately 1949 and 233 points, respectively, indicating that accounting for anti-mAb-
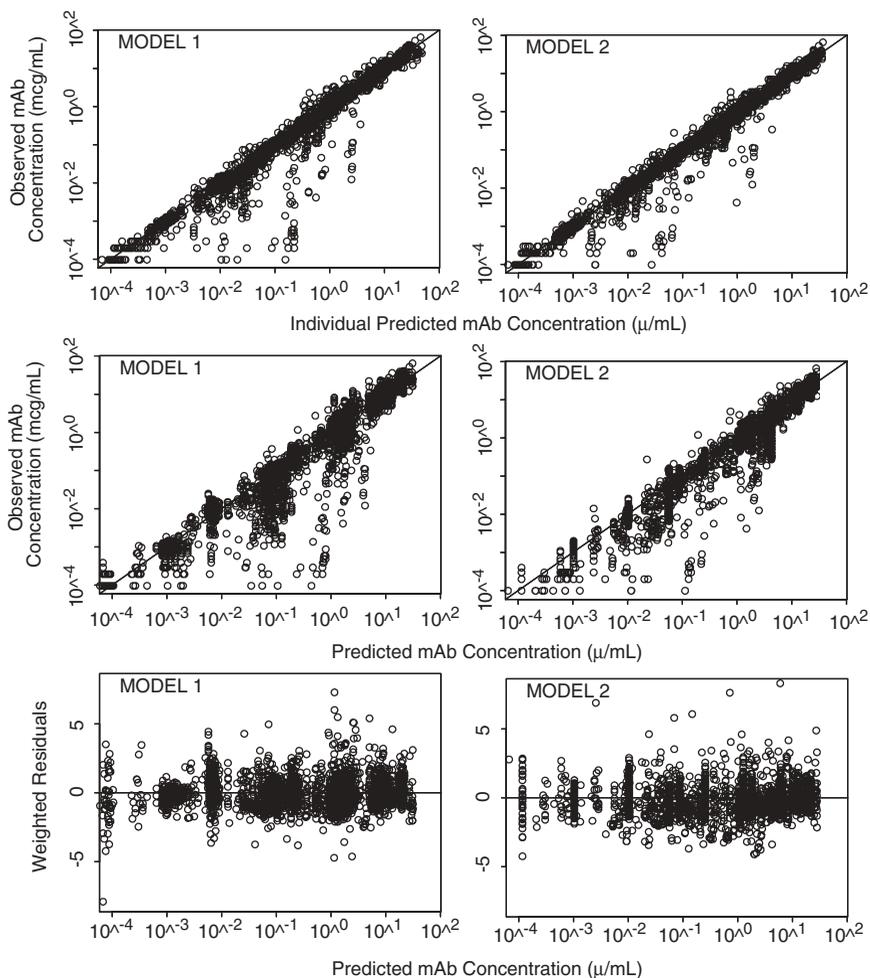
FIGURE 45.4 Goodness-of-fit plots for MODEL 1 and MODEL 2. Plotted data includes all subjects from Phase 1 and Phase 2 studies.

mediated elimination significantly improved the fit to the data. Figure 45.5 includes only those subjects who were anti-mAb positive, and these diagnostic plots suggest that MODEL 2 was able to provide a slightly better fit to anti-mAb-positive subjects; however, both models tended to overpredict the lower concentrations at the end of a dosing interval produced during an anti-mAb response. This is also shown in Figure 45.6, where the lower panels compare the fit of MODEL 1 and MODEL 2 for two subjects in the Phase 1 study who had moderate anti-mAb responses. Both models tended to overpredict the concentrations at later time points during an anti-mAb response, but MODEL 2 provided a better fit to these low concentrations. Also shown in Figure 45.6 in the upper panels are comparisons of the fit of MODEL 1 and MODEL 2 to the nonlinear profile generated in two sample anti-mAb-negative subjects as a result of binding and elimination through the target receptor. The nonlinear profile in these subjects was more accurately fit using MODEL 2 since *CL*
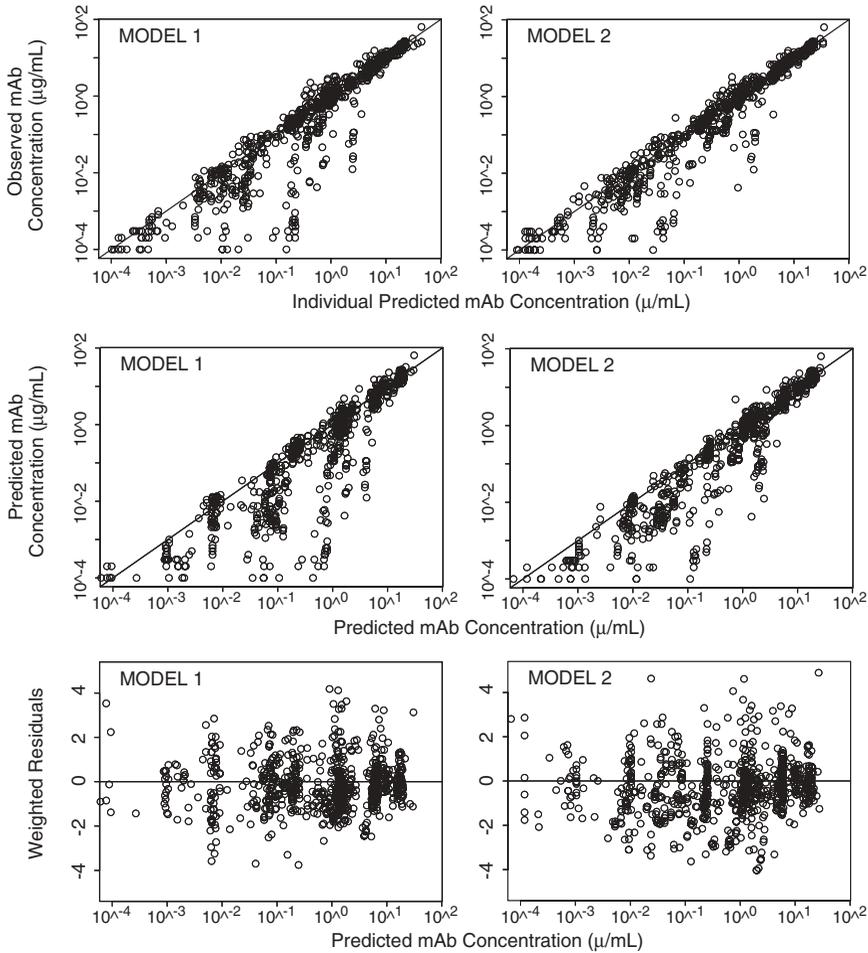
FIGURE 45.5 Goodness-of-fit plots for MODEL 1 and MODEL 2. Plotted data includes only anti-mAb-positive subjects from Phase 1 and Phase 2 studies.

in MODEL 1 was not dependent on concentration of mAb. Over the 0.05–2.0 mg/kg dose range used in the Phase 2 study, MODEL 1 predicted that V would increase from 31.3 to 55.7 mL/kg and CL would decrease from 116 to 21.6 mL/h. MODEL 1 estimated that the anti-mAb-mediated clearance of mAb would be 7.72 mL/h at an anti-mAb $\log_{10}$ titer of 2.5. The estimated value for K_a was approximately three times higher in MODEL 2 compared to MODEL 1. MODEL 2 estimated the maximum rate nonlinear elimination ($V_{\max}$) via anti-mAb to be 0.418 μg/h, with a corresponding K_m of 4.87 μg/mL. The estimated K_e and K_{e1} in MODEL 2 translate to half-lives of approximately 21.9 days and 11.6 h for elimination of free mAb and mAb bound to target receptors, respectively. The estimated value for V was approximately 1.5–2.8 times higher in MODEL 2 compared to MODEL 1. Proportional residual error expressed as a coefficient of variation was 24.4% for MODEL 2 and 29.2% for MODEL 1.
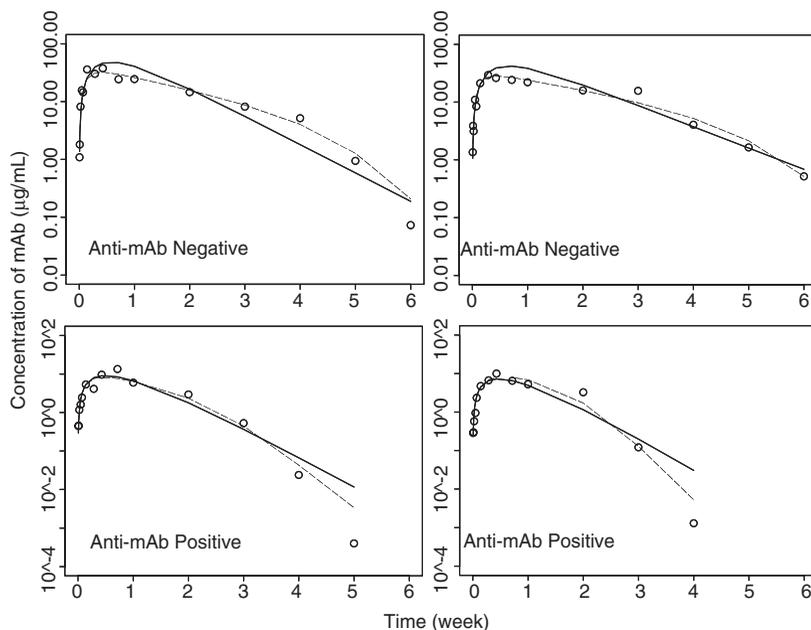
FIGURE 45.6 Comparison of individual model predicted versus observed concentrations in four example subjects in the Phase 1 study. The upper panels show the fit to anti-mAb-negative subjects who had nonlinear elimination resulting only from binding to target receptors. The lower panels show the fit to anti-mAb-positive subjects who had moderate anti-mAb responses, and therefore had two routes of nonlinear elimination. Solid lines represent MODEL 1, dashed lines represent MODEL 2, and open circles represent observed data.

45.3   DISCUSSION

Similar to small molecules, the clinical relevance of PK variability and nonlinearity depend on how strong exposure is correlated to efficacy and safety and the width of the therapeutic index. For therapeutics that have a very wide margin of safety and have an efficacy outcome that can be predicted as accurately with an individual's dose as with an individual's concentration-related exposure, there may be little value in accurately accounting for PK variability and nonlinearity. However, there are many therapeutics where these two conditions are not true. In these cases, it is important to accurately characterize variability and nonlinearity so that one can better understand how changes in dose, patient characteristics, and disease state may affect safety and efficacy through exposure.

The development of antibodies to therapeutic proteins can have direct safety consequences in addition to the impact on pharmacokinetics as addressed in this chapter. Provided the antibody response itself does not have direct safety issues, in some cases it is possible to continue dosing the therapeutic protein in the presence of antibody, provided efficacious levels of the therapeutic protein can be maintained. In this case, it is important to fully understand how the antibody response impacts active concentrations of the therapeutic protein so that it is possible to design dosing regimens that minimize the impact of the antibody response. In the

example discussed in this chapter, the antibody response (anti-mAb) reduced mAb exposure, but this occurred in a nonlinear manner with less negative impact at higher exposures relative to lower exposures. In addition to the influence of anti-mAb, there was an additional nonlinearity observed that was postulated to be the result of mAb binding to the target receptor.

The complexity of the modeling that can be done to understand nonlinearity is highly dependent on the amount and quality of data available. In this example, there were anti-mAb data available that were used to model the influence that the anti-mAb response had on mAb clearance. If anti-mAb data were not available, one could use a mixture model in NONMEM, where the population is assumed to be either anti-mAb positive or anti-mAb negative. This method would improve the fit compared to not accounting for anti-mAb elimination at all; however, the fixed effect of anti-mAb titer has a significant impact on elimination and this method would not be able to account for this fixed effect. In this example, there were also no study data regarding the concentration of the target receptor; however, in many cases this is a biomarker that is available as a measure of target receptor occupancy.

45.3.1 MODEL 1 Advantages, Disadvantages, and Limitations

The main advantage of MODEL 1 is that it is a relatively simple model that uses one of the standard NONMEM models in conjunction with parameter equations that incorporate nonlinearity with respect to dose and anti-mAb titer. This model requires fewer parameters and could be used with sparser data sets. The disadvantage of MODEL 1 is that it is more empirical in nature and does not try to account for the observed nonlinearity on a mechanistic basis. The parameter values obtained from MODEL 1 are questionable from a biological standpoint. The estimated half-life based on apparent CL and V is much lower than would be expected for a mAb, and this is because the clearance in MODEL 1 does not distinguish between the normal linear elimination pathway and elimination via binding to the target receptor. What this means is that MODEL 1 would be unable to accurately predict the clearance of the mAb at doses above or below those used in the modeling. The apparent V estimated with MODEL 1 is at or below the plasma volume; therefore, this volume is not an accurate reflection of the true distribution volume. The apparent V was found to increase with increasing dose. It would be more likely that the apparent V would increase with decreasing dose, since more of the mAb would be bound to the target receptor at lower doses and this would cause less of the mAb to be measured, resulting in an increase in the apparent V.

45.3.2 MODEL 2 Advantages, Disadvantages, and Limitations

The main advantage of MODEL 2 is that it attempts to account for the mechanisms that cause the observed nonlinearity. This should allow MODEL 2 to more accurately predict exposures for doses outside the range currently evaluated or for different dosing schedules, and also to directly account for factors that may impact the number of target receptors or their rate of turnover. The parameter values estimated by MODEL 2 are more realistic from a biological standpoint, with an estimated apparent V of 88.9 mL/kg and half-life of 22 days. The apparent V would

not include binding to target receptors since this is accounted for separately using the equilibrium binding equations. The elimination rate constant (K_{e1}) for target receptor bound mAb could be considered related to the turnover or recycling rate of the target receptor and was found to have a half-life of 12 h. The main disadvantage of MODEL 2 is that its increased complexity and greater number of parameters required a richer data set and also required assumptions about the initial concentration of target binding sites. The binding affinity (K_D) required by MODEL 2 could be reasonably obtained from in vitro binding studies. A limitation of MODEL 2 is that it assumed the number of target binding sites remains constant with treatment and time, whereas in reality it is possible that treatment with the mAb may cause changes in the expression of the target receptor.

45.4 SUMMARY

The example provided in this chapter considers the case of nonlinear pharmacokinetics observed for a protein therapeutic in Phase 1 and Phase 2 studies. The nonlinearity is postulated to be the result of two factors: elimination via anti-mAb antibodies and elimination via binding to target receptors. A comparison is made between fitting the concentration data using an empirical modeling approach versus a mechanistic modeling approach. Both models are able to fit the data reasonably, and the advantages, disadvantages, and limitations of each model are discussed. The best modeling approach to characterize nonlinear pharmacokinetics depends on how much data are available and the intended purpose of the modeling. If the modeling is intended to be used in a predictive manner for future studies that use different dosing regimens and patients with different characteristics, then a more mechanistic approach may lead to more accurate predictions.

REFERENCES

1. T. M. Ludden, Nonlinear pharmacokinetics: clinical implications. *Clin Pharmacokinet* **20**:429–446 (1991).

2. C. A. Van Ginneken and F. G. Russel, Saturable pharmacokinetics in the renal excretion of drugs. *Clin Pharmacokinet* **16**:38–54 (1989).

3. P. Veng-Pedersen, Theorems and implications of a model independent elimination/distribution function decomposition of linear and some nonlinear drug dispositions. I. Derivations and theoretical analysis. *J Pharmacokinet Biopharm* **12**:627–648 (1984).

4. P. Veng-Pederson, Model-independent steady-state plasma level predictions in autonomic nonlinear pharmacokinetics I: Derivation and theoretical analysis. *J Pharm Sci* **73**:761–765 (1984).

5. P. Veng-Pederson, Model-independent method of predicting peak, trough, and mean steady-state levels in multiple intravenous bolus dosing in nonlinear pharmacokinetics. *J Pharm Sci* **72**:1098–1100 (1983).

6. M. T. Smith and T. C. Smith, The unsteady model. An alternative approach to nonlinear pharmacokinetics. *Eur J Clin Pharmacol* **20**:387–398 (1981).

7. C. Pendley, A. Schantz, and C. Wagner, Immunogenicity of therapeutic monoclonal antibodies. *Curr Opin Mol Ther* **5**:172–179 (2003).

8. E. K. Rowinsky, G. H. Schwartz, J. A. Gollob, J. A. Thompson, N. J. Vogelzang, R. Figlin, R. Bukowski, N. Haas, P. Lockbaum, Y. P. Li, R. Arends, K. A. Foon, G. Schwab, and J. Dutcher, Safety, pharmacokinetics, and activity of ABX-EGF, a fully human anti–epidermal growth factor receptor monoclonal antibody in patients with metastatic renal cell cancer. *J Clin Oncol* **22**:3003–3015 (2004).

9. K. R. Lees, H. C. Diener, K. Asplund, and M. Krams, UK-279276, a neutrophil inhibitory glycoprotein, in acute stroke. Tolerability and pharmacokinetics. *Stroke* **34**:1704–1709 (2003).

10. M. Kato, H. Kamiyama, A. Okazaki, K. Kumaki, Y. Kato, and Y. Sugiyama, Mechanism for the nonlinear pharmacokinetics of erythropoietin in rats. *J Pharmacol Exp Ther* **283**:520–527 (1997).

11. D. E. Allison, S. G. Gourlay, E. Koren, R. M. Miller, and J. A. Fox, Pharmacokinetics of rhuMAb CD18, a recombinant humanised monoclonal antibody fragment to CD18, in normal healthy human volunteers. *Biodrugs* **16**:63–70 (2002).

12. R. P. Junghans and C. L. Anderson, The protection receptor for IgG catabolism is the b2-microglobulin-containing neonatal intestinal transport receptor. *Proc Natl Acad Sci* **96**:5512–5516 (1996).

APPENDIX 45.1 MODEL 1 NONMEM CONTROL CODE

```
$PROB FIT OF SINGLE AND MULTIPLE DOSE DATA RUN=001
$INPUT ID TIME APOS DOSE WGT DV MDV AMT EVID LATR

$DATA sdmd.data.final.3.csv IGNORE=C ; Data is in the Excel file.

$SUBROUTINE ADVAN2 TRANS2

;Kineticist: Stuart Friedrich

$PK

  PV=THETA(1)*WGT*DOSE**THETA(5)
  V=PV*EXP(ETA(1))

  IF (LATR .EQ. 0) THEN
    PCL=THETA(2)*DOSE**THETA(6)
  ELSE
    PCL=THETA(2)*DOSE**THETA(6)+THETA(4)*10**LATR
  ENDIF

  CL=PCL*EXP(ETA(2))

  PKA=THETA(3)
  KA=PKA*EXP(ETA(3))

  S2=V
```

```
$ERROR
  IPRED=F
  W=0.001
  IF(F.GT.0) W=F
  IRES=DV-IPRED
  IWRES=IRES/W
  Y=F+W*ERR(1)


$THETA (20,90,150);1 - Volume
$THETA (10,40,100);2 - Linear clearance parameter
$THETA (0.001,0.03,0.1);3 - KA
$THETA (0,0.01);4 - Clearance parameter due to antibody response
$THETA (-0.1);5 - Power exponent for change in V with dose
$THETA (-0.1);6 - Power exponent for change in CL with dose
$OMEGA 0.5;1 - Var of V
$OMEGA 0.5;2 - Var of CL
$OMEGA 0.5;3 - Var of KA
$SIGMA 0.5;PROPORTIONAL ERROR


$EST MAXEVAL=5000 PRINT=5 METH=0 POSTHOC
$COV
$TABLE ID TIME IPRED IWRES APOS LATR DOSE
  FILE=model_sdmd_1_t1.tb NOPRINT ONEHEADER
$TABLE DOSE ID KA V CL ETA1 ETA2 ETA3
  FILE=model_sdmd_1_t2.tb NOPRINT ONEHEADER FIRSTONLY
```

APPENDIX 45.2   MODEL 2 NONMEM CONTROL CODE

Preliminary Evaluated Model

```
$PROB FIT OF SD AND MD DATA RUN=001
$INPUT ID ARM TIME APOS DOSE WGT DV MDV AMT EVID LATR CMT

$DATA data.csv IGNORE=C

$SUBROUTINE ADVAN8 TRANS1 TOL=3

$MODEL
COMP=(COMP1)    COMP=(COMP2,DEFOBS)    COMP=(COMP3)    COMP=(COMP4)
COMP=(COMP5)

$PK

  PV=THETA(11)*WGT
  V=PV*EXP(ETA(11))

  PF1=THETA(1)/150000
  F1=PF1*EXP(ETA(1))
```

```
  PF3=THETA(2)*V
  F3=PF3*EXP(ETA(2))

  PKA=THETA(3)
  KA=PKA*EXP(ETA(3))

  PKF=THETA(4)*1E6
  KF=PKF*EXP(ETA(4))

  PKR=THETA(5)
  KR=PKR*EXP(ETA(5))

  IF (LATR .EQ. 0) THEN
    VMX=0
  ELSE
    VMX=(THETA(6)/150000*10**LATR)*EXP(ETA(6))
  ENDIF

  PKM=THETA(7)/150000
  KM=PKM*EXP(ETA(7))

  PABLG=THETA(8)
  ABLG=PABLG*EXP(ETA(8))

  PKE=THETA(9)
  KE=PKE*EXP(ETA(9))

  PKE1=THETA(10)
  KE1=PKE1*EXP(ETA(10))

  S2=V/150000

$ERROR
  IPRED=F
  W=0.001
  IF(F.GT.0) W=F
  IRES=DV-IPRED
  IWRES=IRES/W
  Y=F+W*ERR(1)

$DES
C2=A(2)/V
A3=A(3)
A5=A(5)
IF (T .LE. ABLG) THEN
  VMAX=0
ELSE
  VMAX=VMX
ENDIF
```

```
DADT(1)=-KA*A(1);MAB ABSORPTION COMPARTMENT

DADT(2)=KA*A(1)-KF*A(3)*A(2)/V+KR*A(4)-KE*A(2)-VMAX*C2/(C2+KM);MAB
FREE COMPARTMENT (CENTRAL)

DADT(3)=-KF*A(3)*A(2)/V+KR*A(5);TARGET SITES FREE COMPARTMENT

DADT(4)=KF*A(3)*A(2)/V-KR*A(4)-KE1*A(4);MAB     BOUND     TO     TARGET
COMPARTMENT

DADT(5)=KF*A(3)*A(2)/V-KR*A(5);TARGET SITES BOUND COMPARTMENT

$THETA (1 FIX);1 - F1
$THETA (1 FIX);2 - F3
$THETA (0,0.03,0.5);3 - KA
$THETA (1 FIX);4 - KF
$THETA (10 FIX);5 - KR
$THETA (0,0.5);6 - VMAX
$THETA (0,1);7 - KM
$THETA (168 FIX);8 - ALAG
$THETA (0,0.002,0.05);9 - KE (LINEAR ELIM)
$THETA (0.01 FIX);10 - KE1 (BINDING ELIM)
$THETA (0,85,150);11 - V/BWT
$OMEGA (0 FIX);1 - F1
$OMEGA (0.5);2 - F3
$OMEGA (0.5 FIX);3 - KA
$OMEGA (0 FIX);4 - KF
$OMEGA (0 FIX);5 - KR
$OMEGA (0.5);6 - VMAX
$OMEGA (0 FIX);7 - KM
$OMEGA (0 FIX);8 - ALAG
$OMEGA (0.5);9 - KE (LINEAR ELIM)
$OMEGA (0.5);10 - KE1 (BINDING ELIM)
$OMEGA (0.5);11 - V/BWT
$SIGMA 0.5;PROPORTIONAL ERROR

$EST MAXEVAL=5000 PRINT=5 METH=0 POSTHOC NOABORT
$COV
```

Final Model

(Note that theta and omega numbering is not the same as in chapter text.)

```
$PROB FIT OF SINGLE AND MULTIPLE DOSE DATA RUN=001
$INPUT ID TIME APOS DOSE WGT DV MDV AMT EVID LATR

$DATA sdmd.data.final.3.csv IGNORE=C ;Data is in the Excel file
```

```
$SUBROUTINE ADVAN6 TRANS1 TOL=3

$MODEL
COMP=(COMP1,DEFDOSE) COMP=(COMP2,DEFOBS)

;Kineticist: Stuart Friedrich
;Notes: Absorption parameters fixed based on fit of single dose
data

$PK

  SIT0=0.000002*EXP(ETA(1))

  F1=THETA(1)/150000

  PKA=THETA(2)
  KA=PKA*EXP(ETA(2))

  PABLG=THETA(3)
  ABLG=PABLG*EXP(ETA(3))

  IF (LATR .EQ. 0) THEN
    VMAX=0
  ELSE
    VMAX=(THETA(4)/150000)*(10**LATR)*EXP(ETA(4))
  ENDIF

  PKM=THETA(5)/150000
  KM=PKM*EXP(ETA(5))

  PKE=THETA(6)
  KE=PKE*EXP(ETA(6))

  PKE1=THETA(7)
  KE1=PKE1*EXP(ETA(7))

  PV=THETA(8)*WGT
  V=PV*EXP(ETA(8))

  S2=V/150000

$ERROR
  IPRED=F
  W=0.001
  IF(F.GT.0) W=F
  IRES=DV-IPRED
  IWRES=IRES/W
  Y=F+W*ERR(1)
```

```
$DES

IF (T .LT. ABLG) THEN
  VMX=0
ELSE
  VMX=VMAX
ENDIF


CBND=A(2)/V*SIT0/(1E-5+A(2)/V)

IF (A(2) .EQ. 0) THEN
  FFRE=1
ELSE
  FFRE=A(2)/(A(2)+CBND*V)
ENDIF


DADT(1)=-KA*A(1)

DADT(2)=KA*A(1)*FFRE-KE*A(2)-VMX*A(2)/V/(A(2)/V+KM)-KE1*CBND*V

$THETA (1 FIX);1 - F FIXED FOR MAB
$THETA (0.0387 FIX);2 - KA
$THETA (168 FIX);3 - ANTIBODY RESPONSE LAG TIME FROM T=0 ONLY
$THETA (0,0.4);4 - VMAX OF ANTIBODY RELATED ELIMINATION
$THETA (0,5);5 - KM OF ANTIBODY RELATED ELIMINATION
$THETA (0.0005,0.0013,0.005);6 - KE (LINEAR ELIM)
$THETA (0.002,0.06,0.2);7 - KE1 (ELIM DUE TO BINDING TO TARGET)
$THETA (70,90,150);8 - V/BWT/F
$OMEGA 0 FIX;1 - VAR OF INTIAL TARGET BINDING SITES AT T=0
$OMEGA 0.0468 FIX;2 - VAR OF KA
$OMEGA 0 FIX;3 - VAR OF ANTIBODY LAG TIME FROM T=0 ONLY
$OMEGA 0 FIX;4 - VAR OF VMAX OF ANTIBODY RELATED ELIMINATION
$OMEGA 0 FIX;5 - VAR OF KM OF ANTIBODY RELATED ELIMINATION
$OMEGA 0.5;6 - VAR OF KE
$OMEGA 0 FIX;7 - VAR OF KE1
$OMEGA 0.5;8 - VAR OF V/BWT/F
$SIGMA 0.5;PROPORTIONAL ERROR

$EST MAXEVAL=5000 PRINT=5 METH=0 POSTHOC
$COV
$TABLE ID TIME IPRED IWRES APOS LATR CBND VMX FFRE DOSE
  FILE=model_sdmd_2_t1.tb NOPRINT ONEHEADER
$TABLE ID KA KM KE KE1 V DOSE
  FILE=model_sdmd_2_t2.tb NOPRINT ONEHEADER FIRSTONLY
```

Development, Evaluation, and Applications of in Vitro/in Vivo Correlations: A Regulatory Perspective

PATRICK J. MARROUM

46.1 INTRODUCTION

With the technological advances in the analytical tools and modeling software available to the pharmaceutical scientist, dissolution testing has been used more and more—both by the industry as well as regulatory agencies as a predictor of differences in bioavailability. When drug release from the formulation and its solubilization are the rate-limiting steps, it is possible to predict the resulting plasma concentration–time profile from its in vitro dissolution. In order to achieve this, there should be a well established relationship between the in vitro dissolution of the drug from the formulation and its in vivo bioavailability.

In this chapter, the various requirements necessary for establishing an in vitro/in vivo correlation (IVIVC) both in terms of in vitro testing and in vivo modeling are presented. The regulatory requirements in terms of validation are discussed. Since the chapter is focused on the regulatory perspective on IVIVC, emphasis is on practical approaches used in drug development and evaluation and less so theoretical aspects of IVIVC. Finally, applications of IVIVC from both an industrial as well as a regulatory perspective are given in terms of obtaining in vivo bioavalability/ bioequivalence waivers and the setting of clinically meaningful dissolution specifications. In this regard, an example is presented in detail to illustrate the various steps in developing and validating an IVIVC.

46.2 LEVELS OF CORRELATION

46.2.1 Level A Correlation

A level A correlation is a point-to-point relationship between in vitro dissolution and the in vivo input rate, as can be seen in Figure 46.1. Such relationships are
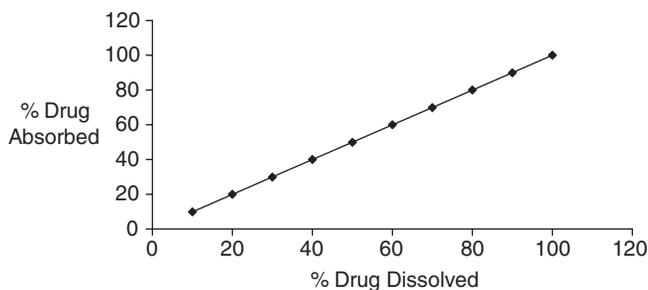
FIGURE 46.1 Level A correlation showing the point-to-point relationship between the fraction of drug absorbed and the fraction of drug dissolved.
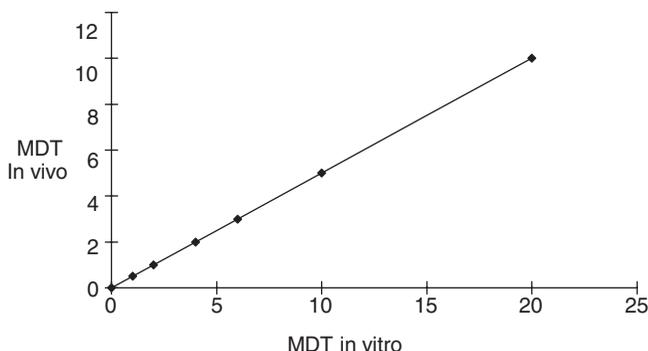
FIGURE 46.2 Level B correlation showing the relationship between the mean in vitro dissolution and the mean in vivo dissolution time.

usually linear, where the in vitro dissolution and the in vivo input curves can be superimposable. Even though nonlinear relationships are uncommon, they can be appropriate since they are useful in predicting the plasma concentration–time profile from in vitro dissolution data (1).

46.2.2 Level B Correlation

In a level B correlation, the mean in vitro dissolution time is compared to either the mean residence time or the mean in vivo dissolution time (Figure 46.2). A level B IVIVC uses the principles of statistical moment analysis. Even though a level B correlation uses all the in vitro and in vivo data, it is not considered a point-to-point correlation. It does not uniquely reflect the actual plasma concentration–time profile because a number of different in vivo profiles will produce similar mean residence times. For this reason, a level B correlation is of little value from a regulatory point of view.

46.2.3 Level C Correlation

A level C correlation establishes a relationship between a dissolution parameter such as the amount of drug dissolved at a certain time and a pharmacokinetic (PK)
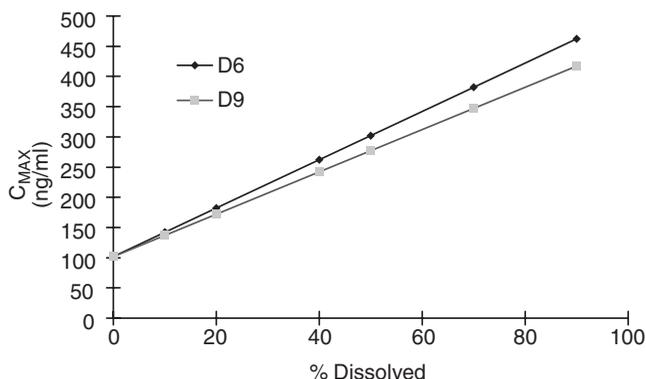
FIGURE 46.3   Level C Correlation showing the relationship between the amount of drug dissolution at a certain time (for example 6 and 9 hours) and the peak plasma concentration.

parameter of interest such as AUC or C_{max} (e.g., see Figure 46.3). Unfortunately, a level C IVIVC does not reflect the complete shape of the plasma concentration–time profile, which is a critical factor in defining the performance of the product. On the other hand, a multiple level C correlation relates one or several PK parameters to the amount of drug dissolved at several time points of the dissolution profile. In general, if one is able to establish a multiple level C correlation, then a level A correlation could be established also and is the preferred correlation to establish.

46.3   DEVELOPMENT OF LEVEL A CORRELATION

46.3.1   In Vivo Considerations

Since a level A correlation is the most useful IVIVC both from a regulatory and formulation development point of view, only the development of a level A IVIVC is discussed in this chapter.

The following points should be taken into consideration when developing an IVIVC:

1. Sine the PK properties of a drug tend to be somewhat different in animals when compared with humans, only human data is considered from a regulatory point of view. This does not preclude the use of animal data in assessing the performance of pilot formulations.
2. The in vivo PK studies should be large enough to characterize adequately the product under study. In general, the larger the variability in the performance of the formulation, the bigger the study should be (2).
3. The preferred study design is the crossover design since it reduces interstudy variability. Parallel studies as well as data obtained across several studies can be utilized to develop the IVIVC.
4. Inclusion of an immediate release reference in the studies facilitates data analysis since it allows one to better estimate the terminal rate constant for

each subject and also enables one to normalize the data to a common reference. The reference product could be an intravenous solution, an aqueous oral solution, or an immediate release product.

5. The studies are usually conducted under fasting conditions. However, if there are any tolerability concerns, the studies could be conducted under fed conditions.

46.3.2 Method

The IVIVC should usually be developed with two or more formulations (preferably three formulations) with different release rates. The process involves the following steps:

- Generate in vitro dissolution profiles using an appropriate dissolution methodology that can discriminate among the various formulations.
- Determine the plasma concentration–time profiles for the tested formulations.
- Obtain the absorption–time profile for these formulations (fraction of drug absorbed versus time). This can be achieved by the use of appropriate deconvolution techniques.
- Plot the in vivo absorption profile or the in vivo dissolution profile against the in vitro dissolution profile to determine whether a relationship exists (e.g., see Figure 46.4).

The method described above is called a two-stage procedure (3). An alternative approach is based on a convolution procedure that attempts to model the relationship between in vitro dissolution and plasma concentrations in a single step. The model predicted plasma concentrations are directly compared to the actual plasma concentrations obtained in the studies (4).

46.3.3 Deconvolution Methods

The most commonly used model-dependent deconvolution methods for estimating the apparent in vivo drug absorption following oral administration are the Wagner–Nelson (5) method and the Loo–Riegelman method (6). These methods depend on mass balance and the fraction of drug absorbed for a one-compartment model is expressed as

$$F_a(t) = \frac{(X_a)_t}{(X_\infty)} = \frac{C + k\int_0^t Ct\, dt}{k\int_0^\infty Ct\, dt} \tag{46.1}$$

where $F_a(t)$ is the fraction of absorbable drug at time t, C is the concentration of drug in the central compartment at time t, and k is the first-order elimination rate constant.

For a two- or three-compartment model, the following equation describes the amount of drug absorbed at time t where V_c is the volume of the central compart-
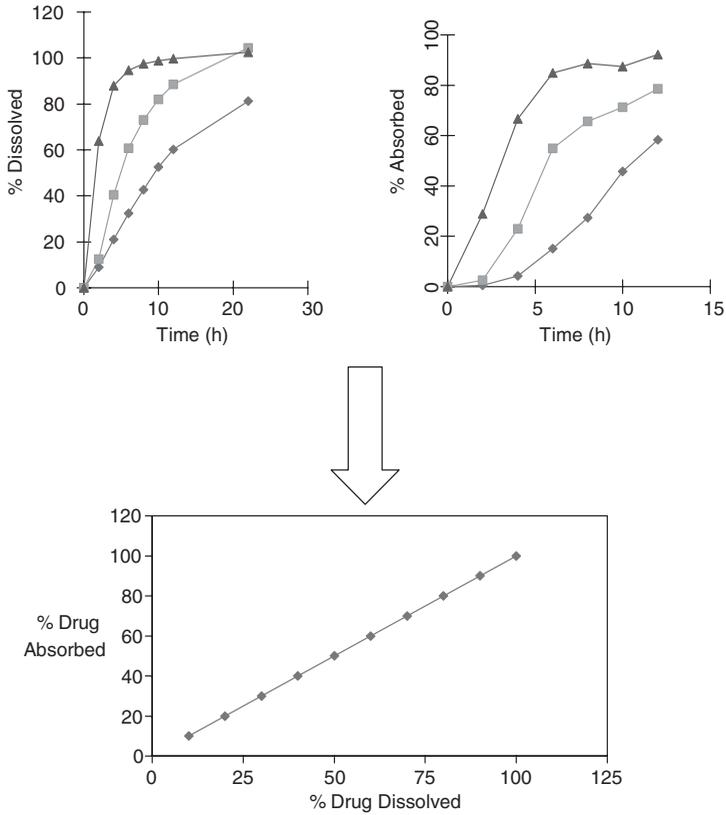
FIGURE 46.4   Development of a level A correlation, where the fraction of drug dissolved at each time is plotted against the corresponding fraction of drug absorbed at the same time. The top left panel represents the dissolution profiles for three different formulations and the top right panel shows the corresponding percent absorbed plots calculated from the respective in vivo absorption profiles. The bottom panel is a synthesis of the information from the two top panels relating the percent absorbed in vivo to the percent dissolved in vitro.

ment, Ct is the plasma concentration at time t, and k_{12}, k_{21}, k_{13}, k_{31} are the inter-compartmental rate constants, and K_{10} is the elimination rate constant from the central compartment (7, 8).

$$\frac{(Xa)_t}{V_c} = C_T + k_{12}\ exp^{-k_{21}\bullet t}\int_0^t Ct\ exp^{-k_{21}\bullet t}dt + k_{13}\ exp^{-k_{31}\bullet t}\int_0^t Ct\ exp^{-k_{31}\bullet t} + k_{10}\int_0^t Ct\ dt$$

(46.2)

46.3.4 Convolution-Based IVIVC

In order to be able to develop an IVIVC using a convolution-based approach, the following assumptions should hold true:

- The in vitro release rate approximates the in vivo absorption rate.
- The PK properties of the drug are linear and time invariant.
- The pharmacokinetics of the drug administered as IV or immediate release (IR) or drug released from the extended release (ER) formulation are indistinguishable. In others words, once a drug molecule released from the IR or ER formulations is absorbed into the systemic circulation, it behaves just like an intravenously administered one.

In addition, plasma concentrations from an IV dose or from administration of IR formulation such as an oral solution or rapidly dissolving oral formulation are needed to estimate the unit impulse function.

If the above conditions are met, then the plasma concentrations are expressed according to the following equation:

$$C(t) = \int_0^t C_\delta(t-u)x'_{\text{rel,vitro}}(u) \, du \qquad (46.3)$$

where $C(t)$ is the plasma concentration at time t, $x_{\text{rel,vitro}}$ is the cumulative amount of drug released in vitro, and x' is the in vitro release rate obtained by taking the first derivative of x. $x_{\text{rel,vitro}}$ can be expressed as any mathematical function that best fits the dissolution profile. Alternatively, x can be estimated by linear interpolation of the mean in vitro dissolution profile. C_δ is the unit impulse response, which is the plasma concentration–time course resulting from the instantaneous in vivo release of a unit amount of a drug (9). C_δ can be obtained by fitting the IR or the IV plasma concentration–time profile to a polyexponential function.

46.4   EVALUATION OF THE PREDICTABILITY OF THE IVIVC

Once an IVIVC has been established, a crucial determination of its applicability is its ability to predict the plasma concentration–time profile accurately and consistently. A relationship between the vitro dissolution and the in vivo absorption rate that is dependent on the release rate of the formulation, as can be seen in Figure 46.5, is usually an indication that a consistent relationship predictive of the in vivo performance does not exist. This is due to the fact that depending on the formulation used one can have a different amount of drug absorbed for the same amount of drug dissolved. On the other hand, a good and consistent relationship would always give you approximately the same slope irrespective of the formulation (whether the slow, fast, or medium formulation is used and whether or not all the data is pooled together). A good illustration of a valid linear level A correlation is presented in Figure 46.6, where the slope of the relationship is the same for each of the individual formulations or for the case where all the formulations are pooled together and treated as one.

Since the IVIVC model is going to be used to predict the plasma concentration–time profile, it is imperative to assess the predictive performance of the model via the assessment of the prediction error of the model. Depending on the intended application of the IVIVC and the therapeutic index of the drug, evaluation of the internal or external predictability may be warranted. Evaluation of internal predictability is based on the data that was used to develop the IVIVC. Evaluation of
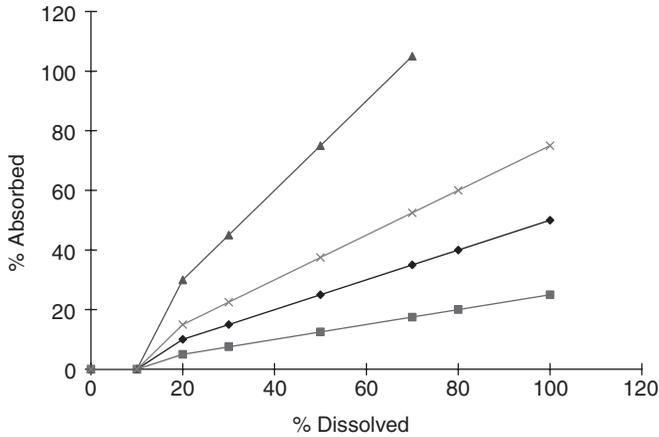
FIGURE 46.5 Poor IVIVC, where the slope of the relationship is dependent on the formulation. Each curve is for a different formulation.
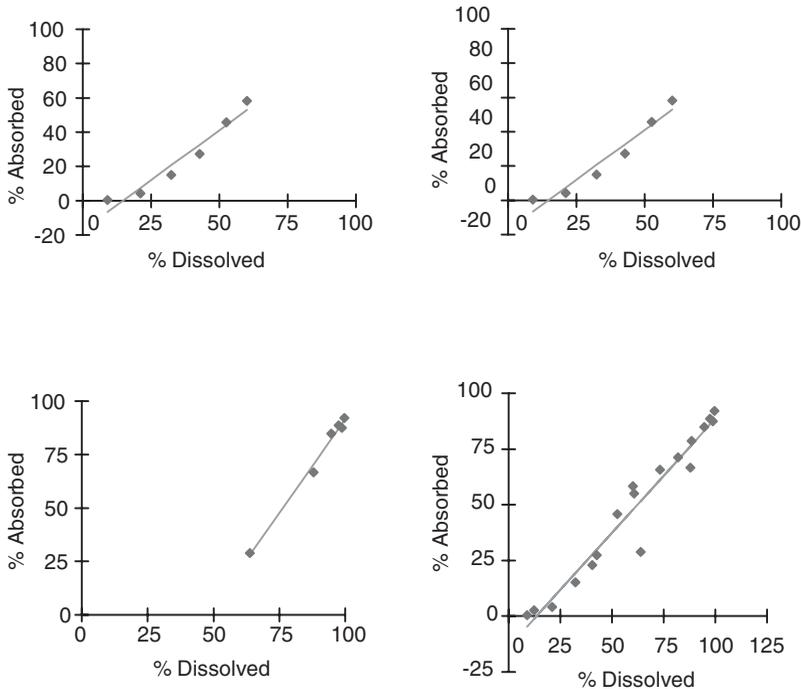
FIGURE 46.6 Predictive IVIVC independent of the release rate, where the slope of the relationship is independent of the formulation used. Each plot represents a different formulation.

external predictability involves additional data sets (see the next paragraph) that were not used in the initial development of the IVIVC.

If the IVIVC for a non-narrow therapeutic index drug was developed with formulations with three or more release rates, the evaluation of the internal predictability would be sufficient to determine its appropriateness.

External predictability is warranted in the following situations:

- The drug is considered to be a narrow therapeutic index drug.
- The internal predictability criteria are not met.
- The IVIVC was developed with two formulations with different release rates.

The data set that is used in the external predictability should ideally be obtained from a formulation with a different release rate. However, it is acceptable to use formulations with similar release rates as those used in the development of the IVIVC. The following represent in decreasing order of preference the types of formulations that can be used to estimate the external prediction errors:

- Formulations with different release rates
- A formulation that was made involving a specific manufacturing change (equipment, process, site, etc.)
- Similar formulations but different lots than the ones used in the IVIVC and the data from a different study than the one used in the development of the IVIVC

46.5 APPROACHES TO THE EVALUATION OF PREDICTABILITY

The most common approach to evaluating the predictability of an IVIVC is depicted in Figure 46.7. The procedure involves the conversion of the in vitro dissolution rate into in vivo absorption rate and then, by the use of convolution methods, a prediction of the plasma concentration–time profile. This is represented as

$$\text{Dissolution} \xrightarrow[\text{IVIC model}]{} \text{Absorption} \xrightarrow[\text{PK parameters}]{} \text{Plasma profile}$$

The area under the concentration–time curve (AUC) and the peak plasma concentration (C_{max}) from the predicted profiles are compared to those obtained from the observed profiles to calculate the percent prediction errors.

The absolute prediction errors are calculated as follows:

$$\left| (\text{Observed} - \text{Predicted})/\text{Observed} \right| \times 100$$

These calculations should be done for each of the formulations used to develop the IVIVC.

For internal predictability, an average absolute prediction error of less than 10% for both AUC and C_{max} establishes the predictive ability of the IVIVC. In addition, the percent error for each formulation should not exceed 15%. If the above criteria are not met, the IVIVC is declared inconclusive and in this case the evaluation of the external predictability of the IVIVC is required.

For external predictability, the percent prediction error should be less than 10% to declare the IVIVC acceptable. A percent prediction error between 10% and 20% is deemed inconclusive, requiring the further evaluation with additional data sets.
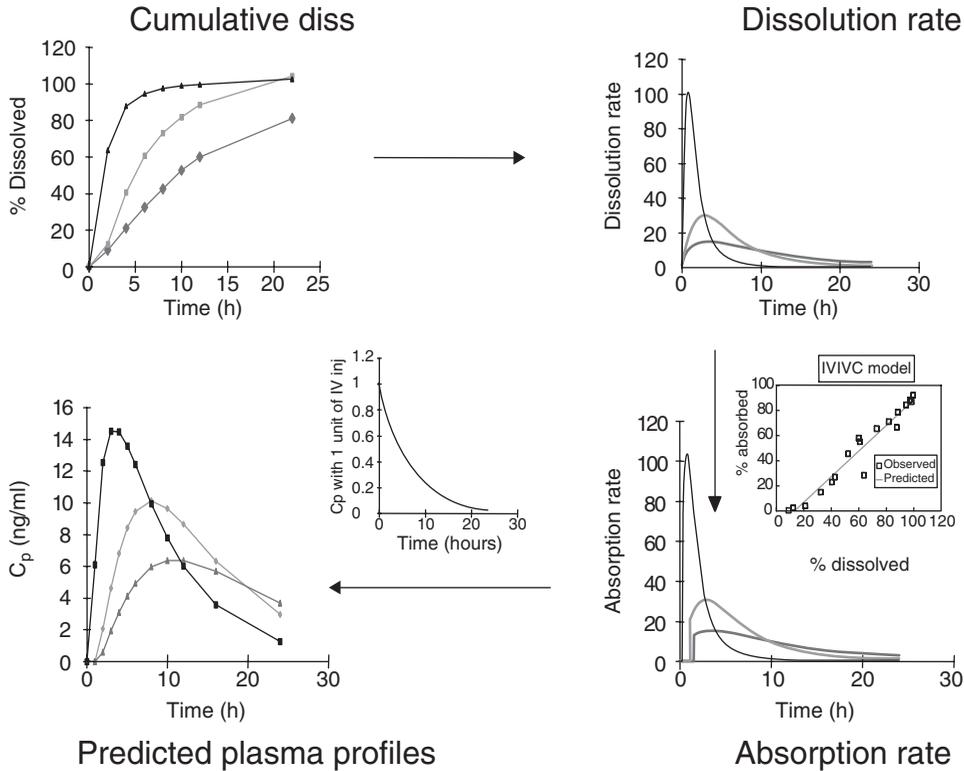
FIGURE 46.7   Most common approach in evaluating the predictability of an IVIVC—the conversion of in vitro dissolution rate into in vivo absorption rate and the prediction of plasma concentration–time profile by the use of deconvolution methods (see Section 46.5).

A percent prediction error greater than 20% indicates that the IVIVC has a poor predictive ability and thus is considered not useful for any application.

Note that the prediction should be made using mean data (mean dissolution profiles as well as population means for the PK parameters) for the following reason: individual dissolution data on the dosage unit that the individual subject was administered is not available. Thus, using average in vitro parameters and individual PK parameters is not appropriate.

Since the purpose of the IVIVC is to predict the performance of yet untested formulations, no individual data will be available for such formulations and therefore a decision as to the appropriateness of the in vivo performance of the formulations is best determined on the average performance of these formulations.

46.6   APPLICATIONS OF IVIVC

46.6.1   In Vivo Bioavailability Waivers

With a predictive IVIVC, in vitro dissolution would not only be a tool to assure the consistent performance of the formulation from lot to lot but would become a

surrogate for the in vivo performance of the drug product. The ability to predict the plasma concentration–time profile from in vitro data will reduce the number of studies required to approve and maintain a drug product on the market, and therefore reduce the regulatory burden on the pharmaceutical industry.

Once an IVIVC has been established, it is possible to waive the requirements for bioavailability/bioequivalence studies. For example, a biowaiver can be granted for a level 3 process change as defined in SUPAC MR, complete removal or replacement of non release controlling excipient, and level 3 changes in the release controlling excipients (10). If the IVIVC is developed with the highest strength, waivers for changes made with the lowest strengths are possible if these strengths are compositionally proportional or qualitatively the same, the in vitro dissolution profiles are similar, and all the strengths have the same release mechanism (11).

However, an IVIVC cannot be used to gain the approval of (a) a new formulation with a different release mechanism, (b) a formulation with a dosage strength higher or lower than the doses that have been shown to be safe and effective in the clinical trials, (c) another sponsor's oral controlled-release product even with the same release mechanism, and (d) a formulation change involving an excipient that will significantly affect drug absorption.

The regultory criteria for granting biowaivers is outlined in the FDA guidance on this topic. Basically, the mean predicted C_{max} and AUC from the respective in vitro dissolution profiles should differ from each other by no more than 20% (see Figure 46.8) (12).

46.6.2 Dissolution Specifications

The IVIVC allows one to shift the dissolution criteria from the in vitro side to the in vivo side. The plasma concentration–time profiles that correspond to the lots that are on the upper and lower limits of the dissolution specifications are predicted. Acceptable dissolution specification limits are limits that do not result in more than 20% difference in AUC and C_{max} (usually ±10% of the target/bio formulation) (13).
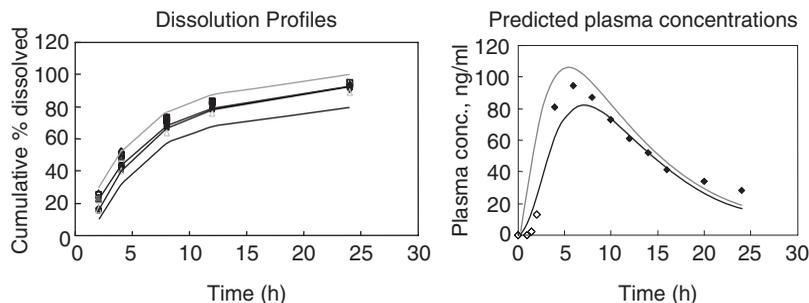
FINAL DISSOLUTION SPECIFICATIONS:
Set such that the predicted Cmax and AUC range NMT 20%

FIGURE 46.8 Regulatory criteria for granting a biowaiver using an IVIVC. The upper of the two profiles is for the test formulation, and the lower of the two profiles is for the reference formulation. NMT = no more than.

Using the IVIVC to choose clinically meaningful specifications provides several advantages in that (a) it will minimize the release of lots that are different in their in vivo performance, thus optimizing the performance of the product; and (b) in certain cases it will allow wider dissolution specifications.

46.7 CASE STUDY

The following example is presented to illustrate the type of study and data analysis that was undertaken to develop a level A correlation for a once-a-day modified release formulation for metopolol.

The in vivo performance of three modified release (MR) formulations with different release rates (fast, medium, and slow formulations) were tested in a four-way crossover single-dose study in healthy volunteers. A fourth other arm of the study included an IR solution of the drug. The individual plasma concentrations are presented in Table 46.1. The mean plasma concentration–time profile for each treatment is shown in Figure 46.9.

TABLE 46.1   Individual Concentrations for the Four-Way Crossover Study

Time (hours)	Plasma Concentrations (ng/mL)						
	Pt1	Pt2	Pt3	Pt5	Pt6	Pt7	Pt9
Solution							
0	ND	ND	ND	ND	ND	ND	ND
0.25	2.51	<1.00	5.67	ND	10.8	3	1.62
0.5	18.4	8.35	13.6	5.51	40.2	35.8	24.3
1	54.3	20.3	54.6	17.7	66.7	46.7	46.4
1.5	76.1	36.9	56.6	37.6	69.5	52.3	52.8
2	84	41.9	56.7	35.9	63.7	48.3	53.2
2.5	78.4	45.9	51.7	42.6	53	41.6	48.4
3	76	51.9	46	39	46.8	36.5	46.4
4	61.3	46.1	34	34.4	33.2	27.9	36.3
6	33.7	33.3	20	21.1	18.3	15.5	23.6
8	22.7	24.1	12.1	15.2	10.8	9.89	14.9
12	7.66	11	2.96	6.54	3.02	3.06	6.16
18	1.86	4.62	<1.00	2.39	1.12	1.25	1.97
24	<1.00	1.74	<1.00	<1.00	<1.00	<1.00	<1.00
Fast Formulation							
0	ND	ND	ND	ND	ND	ND	ND
0.5	7.97	12.3	8.31	1.05	10.8	2.31	7.79
1	64.1	47.5	40.5	6.42	41	31.5	46.9
1.5	124	51.9	83.2	30	74.1	46.6	75.2
2	155	75.5	100	71.7	96	63.5	88.3
3	178	82.1	130	114	120	84	120
4	158	77.4	108	105	107	88.1	107
6	93.9	53.9	67.4	75.5	73.3	48.6	—
8	60.4	36.9	47	51.8	43.3	32.8	49.2
10	38.9	20.9	28.6	32.6	30.9	16.9	27.9

TABLE 46.1 *Continued*

Time (hours)	Plasma Concentrations (ng/mL)						
	Pt1	Pt2	Pt3	Pt5	Pt6	Pt7	Pt9
12	23.9	12.8	18.3	23.8	20.7	9.75	18.7
14	14.5	9.4	12.6	16.7	12.4	6.5	13.3
16	10.9	6.56	8.05	12.7	5.56	4.79	8.77
20	5.93	3.58	4.47	7.11	2.72	2.49	4.83
24	2.77	1.53	2.35	4.62	2.42	<1.00	2.53
30	1.48	<1.00	<1.00	1.84	<1.00	<1.00	1.19
36	<1.00	<1.00	<1.00	1.09	<1.00	<1.00	<1.00
48	<1.00	<1.00	ND	<1.00	ND	ND	<1.00
			Moderate Formulation				
0	ND	ND	ND	ND	ND	ND	ND
0.5	1.98	3.14	3.76	<1.00	8.36	6.4	7.45
1	35.6	26.5	16.8	12.8	33.8	28.3	30.3
1.5	55.1	52.3	24.5	29.7	59	52.8	42.8
2	95	76.4	33.7	43	71.7	57.6	58.1
3	148	100	54.5	78.7	93.7	60	76.9
4	150	106	49.7	87.4	103	54.7	73.5
6	112	99.5	38.1	86.1	77	51.1	71.1
8	86.3	82.8	26	65.9	43.3	34.3	60.4
10	54.2	52.7	16.8	46.9	25.1	25.4	40.5
12	33.6	35.2	9.49	32.1	18.4	14.5	27.3
14	18.9	24.4	5.8	22	11.1	8.35	17.7
16	12.3	17.2	3.65	18	6.78	5.74	10.6
20	6.9	10.5	1.89	9.75	3.34	2.85	5.69
24	3.44	5.81	1.09	4.97	1.86	1.09	—
30	1.51	2.06	<1.00	2.13	—	<1.00	1.28
36	<1.00	1.51	<1.00	1.02	<1.00	ND	<1.00
48	<1.00	<1.00	ND	<1.00	ND	<1.00	<1.00
			Slow Formulation				
0	ND	ND	ND	ND	ND	ND	ND
0.5	2.53	2.61	13.8	<1.00	6.02	8.58	12.4
1	9.92	12.2	25	3.06	19.6	33.8	28.4
1.5	24.3	34.7	47.8	9.2	33.3	34.5	—
2	32	38	62.4	18	42.8	38.8	43
3	45.3	56.8	77.7	47.1	57.7	40.3	72
4	55.8	60.2	85.6	59.9	57.7	42	75
6	53.5	58.1	83.1	80	62.4	37.8	77.1
8	44.7	44.1	69.3	71.5	54.3	28.1	70.5
10	35.6	32.9	46.5	50.4	36.6	16	48.5
12	26.7	25.3	30.2	38.2	26.5	9.49	40.1
14	20.4	18	17.7	29.1	16.3	5.13	28.8
16	14.3	11.8	10.8	21.5	10.2	3.57	20.8
20	9.25	6.43	5.18	13	2.98	1.79	10
24	4.5	3.48	2.8	6.82	1.32	<1.00	5.18
30	1.77	<1.00	1.16	3.6	<1.00	<1.00	1.99
36	<1.00	<1.00	<1.00	2.02	ND	ND	<1.00
48	<1.00	ND	<1.00		ND	ND	ND
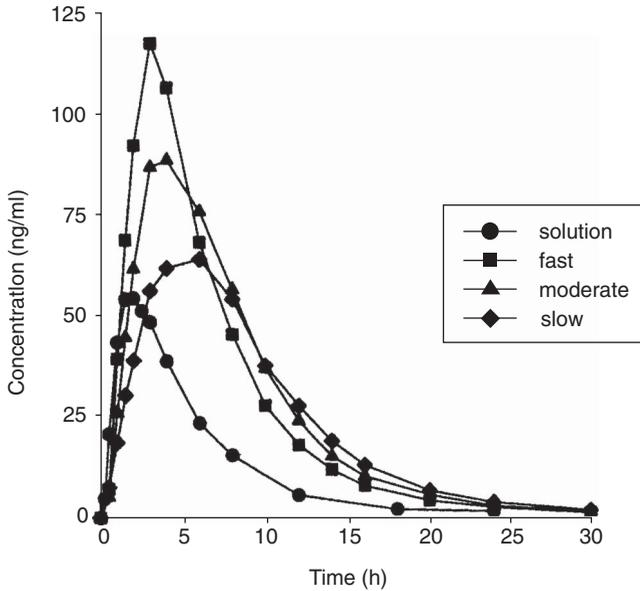
FIGURE 46.9   Mean plasma concentration versus time profile for each treatment.
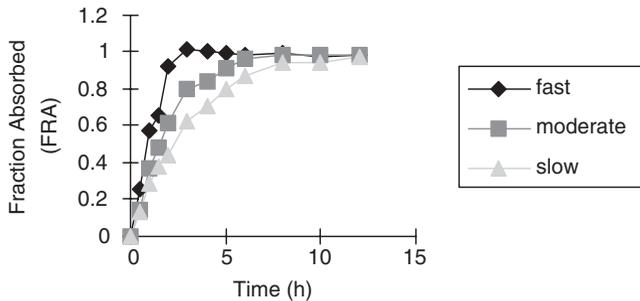
FIGURE 46.10   Mean fraction drug absorbed versus time for each formulation (fast, moderate, and slow) for metoprolol ER study.

The in vitro dissolution profiles were generated under different conditions ranging from pH 1.2 to pH 6.8 at speeds ranging from 50 rpm to 150 rpm using USP apparatus II (rotating paddle) . The terminal rate constant for each subject was determined by linear regression of the linear portion of the log-concentration versus time profile. The peak plasma concentration was the highest observed concentration. The area under the plasma concentration–time curve was determined using the trapezoidal rule. The percent of fraction of drug absorbed from each formulation versus time was determined using numerical deconvolution, where the oral solution derived PK rate constants for each subject were used as the unit impulse function (see Figure 46.10). The numerical deconvolution program PC Decon was used to perform the analysis. Table 46.2 shows the mean fraction of drug dissolved and absorbed used for the fast, medium, and slow formulations.

TABLE 46.2   Mean Fraction of Drug Dissolved and Absorbed Used for the Fast, Medium, and Slow Formulations

Time (hours)	Mean Data	
	Fraction Dissolved	Fraction Absorbed
Formulation: Fast Metoprolol Extended Release		
0	0	0
0.5	0.199	0.254143
1	0.312	0.572286
1.5	0.402	0.658857
2	0.475	0.919286
3	0.611	1.010429
4	0.744	1.005229
5	0.894	0.991857
6	0.971	0.982886
8	0.979	0.992286
10	0.98	0.974286
12	0.979	0.982
Formulation: Moderate Metoprolol Extended Release		
0	0	0
0.5	0.163	0.145857
1	0.253	0.3694
1.5	0.323	0.479143
2	0.386	0.617286
3	0.488	0.801657
4	0.574	0.845
5	0.648	0.914286
6	0.712	0.964414
8	0.816	0.986571
10	0.899	0.981571
12	0.959	0.983
Formulation: Slow Metoprolol Extended Release		
0	0	0
0.5	0.124	0.132586
1	0.19	0.282143
1.5	0.246	0.376571
2	0.293	0.444143
3	0.372	0.620571
4	0.44	0.709286
5	0.502	0.797429
6	0.555	0.873
8	0.646	0.942857
10	0.722	0.948143
12	0.787	0.975571
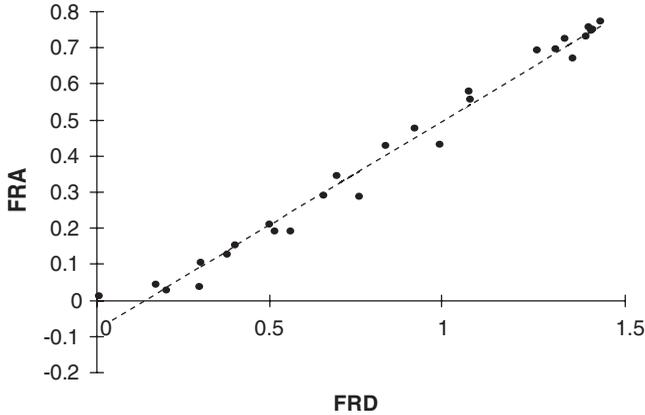
FIGURE 46.11   Relationship between fraction of drug absorbed and fraction dissolved. FRD is fraction of drug dissolved, and FRA is fraction of drug absorbed.

A linear regression analysis was used to examine the relationship between the fraction of drug absorbed and the fraction dissolved. Figure 46.11 shows the relationship between the fraction of drug absorbed versus the fraction of drug dissolved.

The IVIVC model predicted metoprolol plasma concentrations were determined by the following procedure:

- The in vitro dissolution profile corresponding to each formulation was fitted to the corresponding Hill equation to determine the corresponding parameters that describe the shape of the dissolution profile:

$$\% \, Dissolved = \frac{D_{max}T^{\gamma}}{D_{50}^{\gamma} + T^{\gamma}} \tag{46.4}$$

  where D_{max} is the maximum amount of drug dissolved, D_{50} is the time required for 50% of the drug to dissolve, γ is the sigmoidicity factor, and T is time.
- The corresponding dissolution rate was obtained by taking the first derivative of the corresponding Hill equation for the respective formulation.
- Using the IVIVC relationship, the in vivo dissolution rate for each formulation was determined.
- The predicted plasma concentration–time profile for each formulation was obtained by convolution of the in vivo dissolution rate and the PK model describing the oral solution data. A one-compartment model with a mean elimination rate constant of $0.29 \, h^{-1}$ and a volume of distribution of $5.9 \, L/kg$ was adequate to describe the mean oral solution profile. The convolution was accomplished on a spreadsheet in Lotus 1-2-3.
- Table 46.3 show the percent prediction error for each of the formulations used to develop the correlation for both C_{max} and AUC (14).

It can be seen from the results that this IVIVC is considered to be predictive because it met the internal validation criteria in that the mean prediction errors for both C_{max} and AUC were below 10% with none of the individual ones exceeding 15%.

TABLE 46.3   The Percent Prediction Error for Each of the Formulations Used to Develop the Correlation for Both C_{max} and AUC

Formulation	C_{max}	AUC
Slow	−5.67	−0.76
Moderate	−0.85	5.22
Fast	3.97	4.52
Average	3.50	3.50

46.8   SUMMARY

The establishment of a predictive relationship between in vitro dissolution and the in vivo bioavailability of a modified release formulation is discussed. Establishing an IVIVC would decrease the number of studies required to approve and maintain a product on the drug market. A predictive level A correlation would enable the in vitro dissolution to become a surrogate for the in vivo performance of the drug product. This is one of the instances where regulatory decisions are made based on modeling and simulation data. The ability to describe both the in vitro and in vivo performance with well established mathematical models and the availability of different software that are able not only to fit the data but predict the resulting plasma concentration–time profiles should make the development, evaluation, and applications of IVIVC a routine endeavor in the development of modified release formulations.

REFERENCES

1. US Pharmacopeia 23. General Information/In vitro in vivo Evaluation of Dosage Forms. <1088>, 1995, pp. 1924–1929.

2. *FDA Guidance for Industry: Extended Release Oral Dosage Forms: Development, Evaluation and Applications of in Vitro/in Vivo Correlations*, September 1997.

3. P. J. Marroum, *Development of in Vivo in Vitro Correlations, SUPAC-MR/IVIVC Guidance FDA Training Program Manual*. Center for Drug Evaluation and Research, Food and Drug Administration, July 1997, pp. 62–76.

4. W. R. Gillespie, Convolution—based approaches for in vivo in vitro correlation modeling, in vitro in vivo correlations. *Adv Exp Med Biol* **423**:53–65 (1997).

5. J. G. Wagner and C. Metzler, Estimation of rate constants for absorption and elimination from blood concentration data. *J Pharm Sci* **56**:658–659 (1967).

6. L. C. K. Loo and S. Riegelman, Anew method for calculating intrinsic absorption rates of drugs. *J Pharm Sci* **57**:918–928 (1968).

7. F. Langenbucher, Numerical convolution/deconvolution as a tool for correlating in vitro and in vivo drug availability. *Pharm Ind* **44**(11):1166–1171 (1982).

8. F. Langenbucher, Improved understanding of convolution algorithms correlating body response with drug input. *Pharm Ind* **44**:1275–1278 (1982).

9. W. R. Gillespie, Modeling strategies for in vivo in vitro correlations, in *Scientific Foundations for Regulating Drug Product Quality*, G. Amidon, J. R. Robinson, and R. L. Williams (Eds.). AAPS Press, Alexandria, VA, 1997, pp. 275–292.

10. R. Uppoor, *Evaluation of Predictability of in Vivo in Vitro Correlations, SUPAC-MR/ IVIVC Guidance FDA Training Program Manual*. Center for Drug Evaluation and Research, Food and Drug Administration, July 1997, pp. 97–110.

11. P. J. Marroum, Regulatory examples: dissolution specifications and bioequivalence product standards, in *Scientific Foundations for Regulating Drug Product Quality*, G. Amidon, J. R. Robinson, and R. L. Williams (Eds.). AAPS Press, Alexandria, VA, 1997, pp. 305–319.

12. H. Malinowski, P. J. Marroum, and R. Uppoor, FDA, Guidance for Industry: extended release oral dosage forms: development, evaluation and applications of in vitro/in vivo correlations. *Diss Tech* **4**:23–32 (1997).

13. P. J. Marroum, Role of in vivo-in vitro correlations in setting dissolution specifications. *Am Pharm Rev* **2**:39–42 (1999).

14. N. D. Eddington, P. Marroum, R. Uppoor, A. Hussain, and L. Augsburger, Development and internal validation of an in vitro–in vivo correlation for a hydrophilic metoprolol tartrate extended release tablet formulation. *Pharm Res* **15**:466–473 (1998).

The Confluence of Pharmacometric Knowledge Discovery and Creation in the Characterization of Drug Safety

HUI-MAY CHU and ENE I. ETTE

47.1 INTRODUCTION

The success rate of new chemical entities (NCEs) is anything but stellar (1). In 1987 the cost of bringing a new drug into the market was $237 million as opposed to $802 million in 2000 (2). By the end of 1999, 21% of the NCEs with investigational new drug (IND) applications filed from 1981 to 1992 had been approved for marketing in the United States (3). Of those that failed in the period from 1987 to 1992, 38% of the NCEs failed because of efficacy (activity too weak or lack of efficacy), 34% on economics (commercial market too limited, or insufficient return on investment), 20% because of safety (human or animal toxicity), and the rest for nonspecific reasons (3). What is becoming increasingly clear is that traditional drug development approaches are unlikely to succeed in the future given the economics of drug development—a low probability of success coupled with increasing product development times means decreased sales time after market launch and lower return on investment for pharmaceutical companies.

To speed drug development, sophisticated new technologies and approaches in the discovery and design of new drugs are replacing the traditional methods of discovery. Increasingly, however, a pharmacometrically guided approach is being applied to drug development. The need to get the most knowledge from every drug development study that is performed cannot be overemphasized in this day and age of spiraling drug development cost. This need has led to the development of pharmacokinetic/pharmacodynamic (PK/PD) knowledge discovery (4) (see also Chapter 14 of this text) and creation (5, 6) (see also Chapter 32 of this text). Understanding the pharmacodynamics of the drug in addition to its pharmacokinetics can lead to a minimization of drug adverse effects. Pharmacodynamics is a broad term, intended to include all of pharmacological actions, pathophysiological effects, and therapeutic responses, both beneficial and adverse, of an active ingredient,

therapeutic moiety, and/or its metabolite(s) on various systems of the body from cellular effects to clinical outcomes (7). Using pharmacometric knowledge discovery and creation approaches could enable the characterization of an unexplored region of the response surface (6), in terms of not only efficacy but also safety.

PK/PD knowledge discovery is the nontrivial process identifying valid, novel, potentially useful, and ultimately understandable patterns in data by characterizing data structure by means of a model (4). PK/PD knowledge creation, on the other hand, is the process of building upon current understanding of data that is already acquired by generating more data (information) that can be translated into knowledge. It entails the use of (valid) models to synthesize data, estimate inestimable uncertainty, or supplement data for further knowledge acquisition (6). Data synthesis is performed when available knowledge about the drug is used to simulate a clinical trial to explore study outcome when various controllable and uncontrollable factors are varied. This is a knowledge creation process because the objective is to obtain knowledge about the unknown (i.e., unexplored region of the response surface) using valid models. A detailed discussion of data synthesis, estimating inestimable uncertainty, and data supplementation as PK/PD knowledge creation approaches is given by Ette and Chu in Chapter 32 of this book and by Williams et al. (5) and by Ette et al. (4, 6). Data supplementation deals with the use of models on available data to generate supplemental data that will be used to characterize a targeted unexplored segment of the response surface. The assumptions about models to be used in data supplementation are not as stringent as those required for data synthesis. That is, the use of predictive models is not an absolute necessity.

The challenge of clinical drug development is to do the utmost to extract hidden knowledge from clinical trial data and to be able to use that knowledge to plan the next set of trials. Thus, using the understanding of the past/present to gain an understanding of the future, and using present knowledge to learn about what "could have been" are critical to knowledge-driven drug development. This is the setting for the confluence of PK/PD knowledge discovery and creation. Specifically, how do we address the issue of drug safety in a just concluded study at doses not studied? In this chapter an approach to finding a solution to the above question is presented.

In the sections that follow pharmacometric knowledge discovery techniques, the confluence of pharmacometrics knowledge discovery and creation, and an application example to demonstrate how the above question can be addressed in a systematic manner are discussed.

47.2   PHARMACOMETRIC KNOWLEDGE DISCOVERY TECHNIQUES

Techniques such as visualization, generalized additive modeling (GAM), tree-based modeling (TBM), and clustering used in pharmacometric (PM) knowledge discovery are discussed in Chapter 14 of this book and by others (8–11). The reader is referred to that chapter to familiarize himself/herself with these techniques. In addition, a new data structure revelation technique not discussed in Chapter 19 is introduced in this chapter together with a proposed new metric for characterizing adverse events (AEs). This is the percentile division technique for revealing

structure in a data set. In addition, the bootstrap, which is discussed in detail in Chapters 8 and 15 of this book, is discussed briefly in this section.

47.2.1 Percentile Division

Percentile division is a systematic approach to finding a specific value of a covariate that can split data into subgroups to maximize the probability structure in revealing explanatory variables that can be used as predictors of the response variable in a data set. The response variable could be binary, categorical, or continuous. In a data set with a binary outcome variable, for instance, the procedure would be as follows:

Step 1. This involves labeling subjects into two groups based on the distribution of the baseline values of an important covariate, based on some prior information, from all subjects using a given percentile as the division point. Group 1, for instance, would consist of subjects that have baseline values less than the 10th percentile of the baseline values and the rest of the subjects will be denoted as group 2.

Step 2. This involves fitting a statistical model to incorporate the grouping/indicator variable—0 for group 1, and 1 for group 2. The statistical model could be a linear model (LM), generalized linear model (GLM), or generalized additive model (GAM). This is done to select covariates useful for explaining the variability in response. Steps 1 and 2 are repeated for different division cutoff points, that is, in 10 percentile increments up to the 90th percentile. Subsequently, all model fittings are performed and comparisons of modeling results made to provide insight as to where the cutoff point should be that maximally reveals the hidden structure in the data set (see Appendix 47.1).

47.2.2 Percentage of Duration Above a Certain Predefined Grade of an Adverse Event: A New Metric for Characterizing a Safety Outcome

Generally, there are predefined absolute thresholds (e.g., grade 3 or 4 adverse event) before a safety concern can be raised during a clinical study. The typical practice is to track the safety response profile during the course of the study to see if there is a rise above a threshold value at any particular time. Thus, the criterion used for analysis is any occurrence of a particular adverse event during the study. However, few consider this to be inadequate for tracking the time course of an adverse event such as the elevation of a safety biomarker. Figure 47.1 shows two scenarios for two subjects that could be said to have the same grade of AE. However, there is profound impact in terms of duration and severity of the AE for the subject in the lower panel of Figure 47.1 than the subject in the upper panel. By the "any occurrence" definition, both are AE "responders." Which of the subjects would likely suffer more because of the AE? In order to answer this question a robust new metric is proposed that takes into account the time course of a safety biomarker and the duration of the biomarker above a predefined threshold. The metric is the area under the curve above the threshold (AUAT) for an AE grade. The use of this metric would facilitate the evaluation of severity and duration of the safety biomarker above the selected threshold. Thus, from Figure 47.1 it is obvious
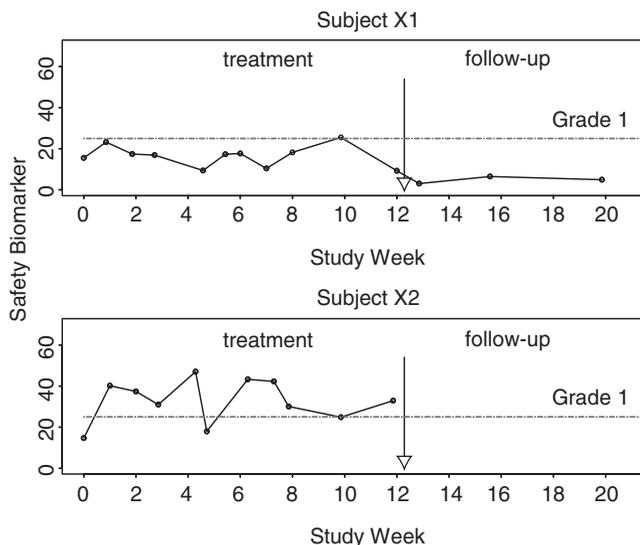
FIGURE 47.1   Example of the evaluation of severity and duration of the safety biomarker above a threshold value for two subjects. Subject X1 had less than 2 days above grade 1 AE, while subject X2 had more than 10 weeks above the grade.

that the subject in the upper panel who spent less than 2 days above the grade 1 AE threshold is affected less by the AE than the subject that spent more than 10 weeks. [See Appendix 47.2 for a sample S-Plus code that can be used to generate this metric.]

47.2.3   The Bootstrap Technology

The principle of the bootstrap is to repeatedly generate pseudosamples distributed according to the same distribution as the original sample (12). The original data set consists of an independent and identically distributed (i.i.d.) sample of size N from an unknown probability distribution. Original distribution, though unknown, may be replaced by the empirical distribution of the sample. Readers are referred to Chapter 15 of this text for an in-depth discussion of the bootstrap. For the purposes of this chapter, as discussed in the application example, the bootstrap was used to create 500 replicate pharmacokinetic data sets to enable regression modeling and prediction of exposures for doses that were not studied in previous trials, herein referred to as target doses.

47.3   THE CONFLUENCE OF PHARMACOMETRIC KNOWLEDGE DISCOVERY AND CREATION

A confluence is the merger or meeting of two or more objects (or subject matters) that seem to inseparably bind their respective forces or attributes into a point of junction. The point of junction of pharmacometric knowledge discovery and creation is in attempting to gain knowledge and understanding of the response

surface, especially an unexplored region. To be able to determine what the effect of unstudied doses in a just completed trial would have been, a combination of knowledge discovery and creation methodologies were implemented on a pooled data set from a few studies.

47.4   APPLICATION

47.4.1   Objective

The purpose of the analysis was to discover knowledge of the underlying relationship between drug administration and safety (i.e., safety biomarker elevation), and to create knowledge about safety outcome with different dosing regimens (i.e., targeted dosages and dosing regimens) that were not studied in a recently completed study.

47.4.2   Data

Data from three studies were pooled and analyzed. Study 1 was a 7 day healthy volunteer study with two dosing regimens (placebo, 150 mg TID, 450 mg BID); study 2 was a 14 day healthy volunteer study with four dosing regimens (placebo, 75 mg BID, 150 mg BID, and 300 mg BID); and study 3 was a 12 week patient study with three dosing regimens (placebo, 150 mg BID, and 300 mg BID).

47.4.3   Methodology

By pooling data from the three studies, a larger data set was available for knowledge discovery and to create knowledge about the targeted doses. However, due to heterogeneity in the populations studied (i.e., healthy subjects vs. diseased subjects) and the nature of the studies—different dosing regimens and different treatment periods—knowledge discovery challenges had to be overcome and assumption validation tested before any inference or extrapolation to an unexplored region of the response surface involving the higher doses could be made.

In performing pharmacometric knowledge discovery, the following questions had to be addressed before proceeding to perform pharmacometric knowledge creation that would address the objective of the investigation.

1. Are the pharmacokinetics of the drug similar across studies and in the different study populations?
2. Are there any covariates that influence the pharmacodynamics (i.e., the safety) outcome?
3. Do treatment durations have any influence on safety biomarker profiles?

47.4.3.1   *Pharmacometric Knowledge Discovery and Creation*
To be able to answer the above questions and thereby address the objective of the investigation, the pharmacometric (PM) knowledge discovery and creation proceeded in three stages. In stage 1, PM knowledge discovery of the relationship between biomarker and exposure was performed with data from the studies

performed in healthy subjects. In stage 2, the procedure was repeated with data from a study conducted in a diseased subject population, and in stage 3 data from studies used in stages 1 and 2 were combined for PM knowledge discovery and creation. All graphical and statistical analyses were performed using S-Plus Version 6.02 (Insightful, Seattle, WA).

Stage 1 Pharmacokinetic (PK) data from the healthy subject studies (studies 1 and 2) were analyzed using the statistical moments analysis approach. From the results of the analysis, peak concentration (C_{max}) and area under the plasma concentration curve (AUC) were selected for exploring the relationship between exposure and safety data (biomarker elevation).

Exploratory data analysis (EDA), involving the use of data visualization, and tree-based modeling (TBM) were performed to reveal any relationship between response (biomarker level) and the individual PK parameters, biomarker baseline values, and subject demographics. Initial data visualization was performed using "trajectory" and pairs plots. Figure 47.2 is a trajectory plot showing the direction of change in biomarker levels in subjects. This would tend to suggest that drug treatment might have caused a change in biomarker levels. However, the picture is not so clear when the biomarker levels are related to exposure parameters (Figure 47.3). Further structure revelation analysis was performed using TBM. Figure 47.4 illustrates the result obtained with the TBM approach.

C_{max} and safety biomarker baseline values were two informative predictors that could be used to explain the response (safety biomarker elevation) based on the first two nodes of the tree derived from the TBM approach. However, the usefulness of further divisions was limited. The TBM approach is useful as an initial exploratory procedure, but its usefulness is enhanced if the derived model is validated.
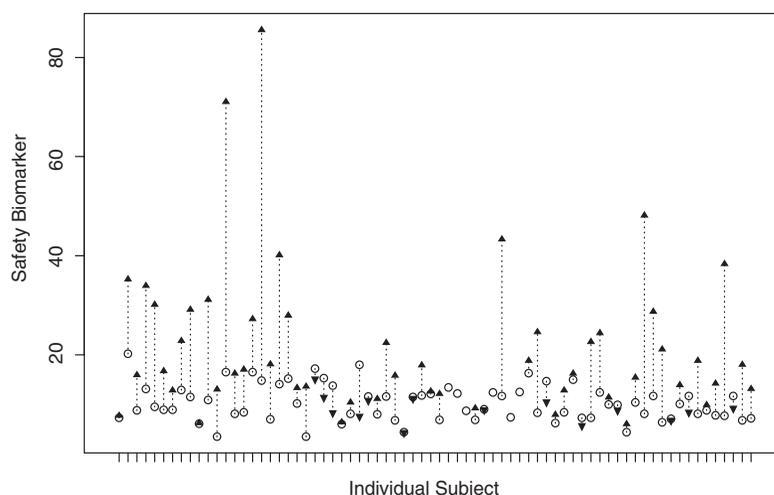
FIGURE 47.2   A trajectory plot showing the direction of change of a safety biomarker for individual subjects. Open circles represent the baseline safety biomarker values, filled triangles are the safety biomarker values at end of treatment, and the dotted lines show the trends from baseline values of the safety biomarker.
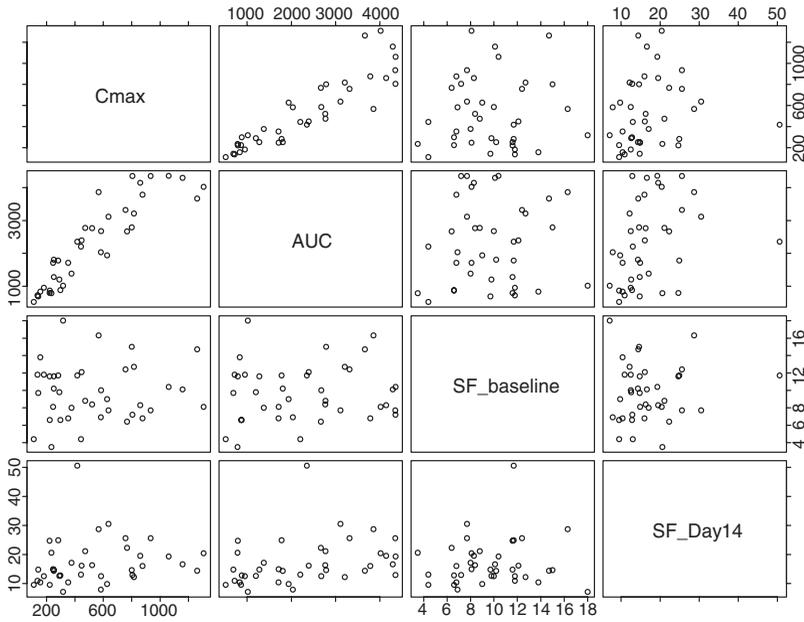
FIGURE 47.3   A pairs plot showing the relationship between covariates (C_{max}, AUC, baseline safety biomarker values (SF_baseline), and safety biomarker level on study day 14 (SF_Day14)).
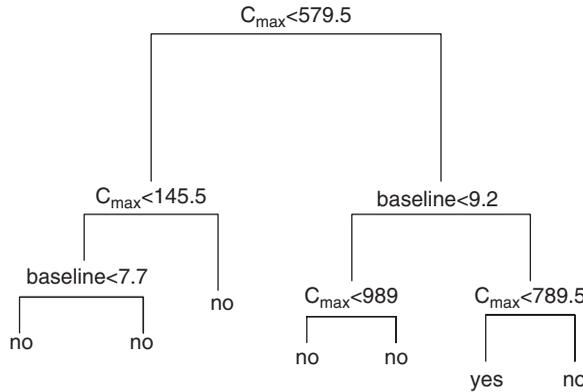
FIGURE 47.4   A classification tree showing the relationship between the occurrence of an AE (yes or no) and significant predictor covariates (C_{max}, safety biomarker baseline values).

TBM was not validated since the intent was to use it as an initial exploratory data analysis screen.

Figure 47.5 is a two-dimensional (2D) bubble plot that illustrates another perspective of examining the relationship between the predictor variables and biomarker response. In essence, the figure is a 2D representation of three-dimensional data. The baseline biomarker values are on the y-axis, C_{max} is on the x-axis, and
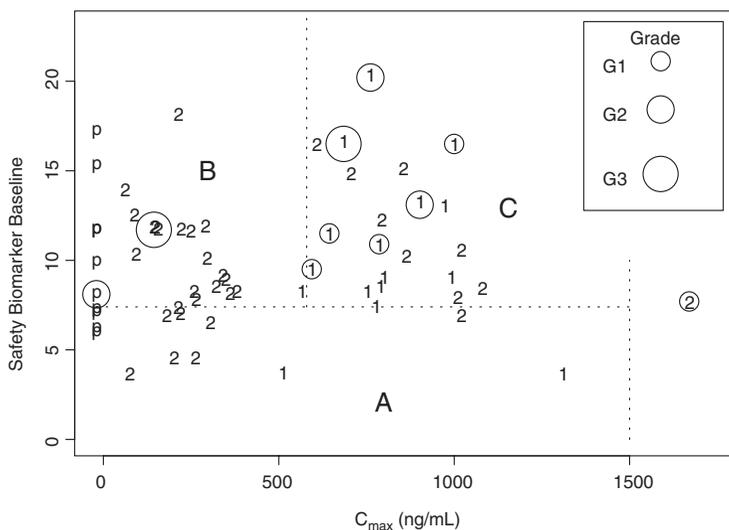
FIGURE 47.5 A two-dimensional (2D) bubble plot used to examine the relationship between the exposure metric (C_{max}) and safety biomarker response. Active treatment subjects from studies 1 and 2 are denoted by 1s and 2s, respectively, and p's are placebo subjects from both studies. The bubbles (open circles) indicate the safety biomarker responses; and the severity of the response is depicted by the size of the circles.

coming off plane is the bubble (with varying sizes) representing the degree of biomarker elevation in grades (i.e., 1, 2, or 3). Placebo treatment is denoted "p" and active treatments are denoted by "1" and "2" from studies 1 and 2, respectively. The dotted lines roughly partition the response surface into three different zones—A, B, and C. A indicates that when baseline values were low, there is no biomarker elevation. B shows that when the C_{max} is low, the treatment subjects have a rate of AE occurrence similar to the "natural occurrence" rate represented by the placebo group. C suggests that when both C_{max} and biomarker baseline values are above certain cutoff values, there appears to be some exposure–response relationship.

Stage 2   The PK data was analyzed using the noncompartmental approach as was done in stage 1. Thereafter, the percentile division approach was used to discern structure in the safety biomarker data. It was observed that diseased subjects could be categorized into two subgroups (i.e., high and low baselines) based on biomarker baseline values. The 50th percentile of the safety biomarker baseline values was found to be the cutoff point for group separation. The safety data from the two subgroups were related to exposure in an additive generalized logistic regression model. Figure 47.6 shows the exposure–response relationship for grade 1 adverse event after the baseline grouping criterion was incorporated into the logistic model. AUC was found to be the best predictor of the safety biomarker elevation. Although the biomarker was measured as a continuous variable, the modeling was done using the categorical grading system used in the clinic. Although converting a continuous variable into an ordered categorical variable results in a loss of information, this was
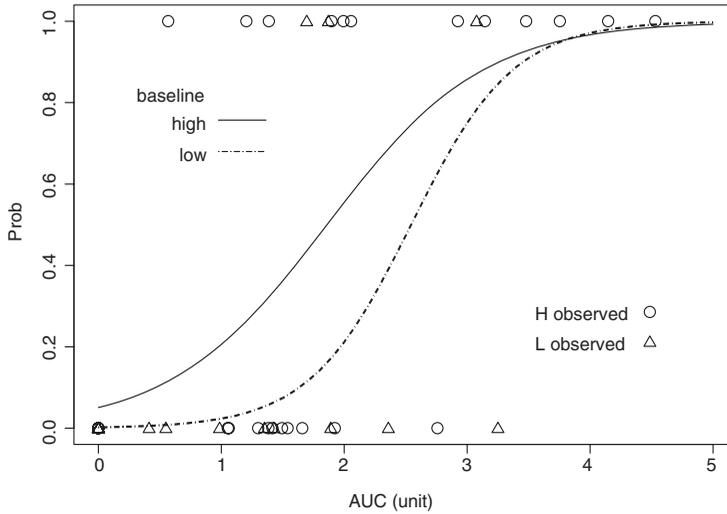
FIGURE 47.6   The exposure–response relationship showing the probability of having grade 1 AE after incorporating the baseline grouping criterion (see text).

done for the sole purpose of facilitating communication with clinicians who always use AE score (an ordered categorical variable) as a measure for monitoring changes in the safety biomarker in practice. Appendix 47.3 contains a sample code that can be used for transforming a continuous variable into a categorial variable.

Figure 47.7 is an example of the characterization of the relationship between AUC and AE using a logistic regression model for subjects with a high biomarker baseline value. The figure shows the probability of a range of occurrences of duration (i.e., from any occurrence to 50% duration) above the threshold for grade 1 AE. Increasing exposure is associated with increasing probability of having the AE, with AUCs of 5 units being associated with 65% probability of having at least 50% duration above the grade 1 AE threshold.

Stage 3

PHARMACOKINETICS   The area under the plasma concentration–time curve (AUC) was identified, in a preliminary analysis, as the important exposure covariate that was predictive of the safety biomarker outcome. Consequently, it became necessary to compare the distributions of AUC values across studies and dosage regimens. Figure 47.8 illustrates distributions of the exposure parameter AUC across studies. It is evident that AUC values are higher in diseased subjects than in healthy volunteer subjects at the same dose level. To adjust for the difference between the two subpopulations, an indicator function was introduced in a first-order regression model to better characterize the dose–exposure data. Let y be the response variable (i.e., AUC), x is a predictor variable, β is the regression coefficient on x, and ε is the error term, which is normally distributed with a mean of zero and variance σ^2. Thus,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 I_i + \varepsilon_i \tag{47.1}$$
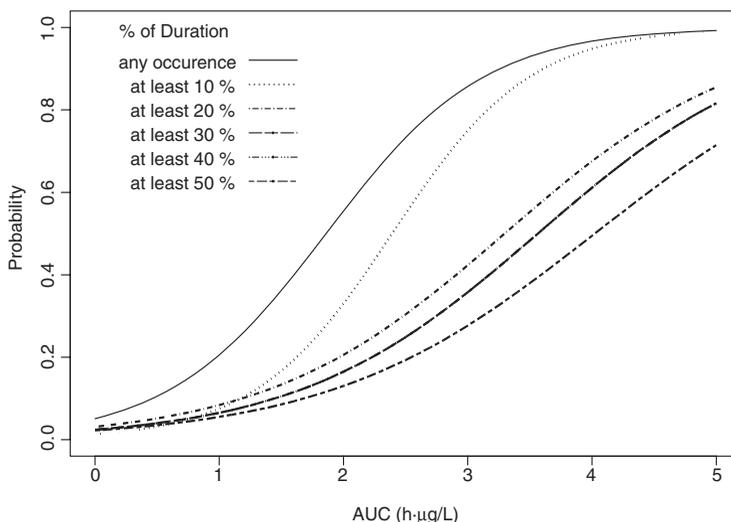
FIGURE 47.7 The exposure–response relationship for the high baseline group showing probability of having grade 1 AE given different percentages of study duration (study 3 only—diseased subjects).
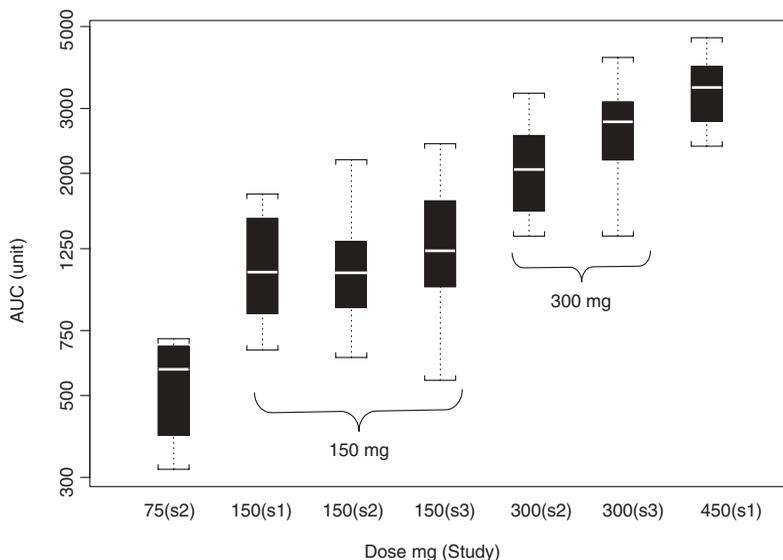
FIGURE 47.8 Distribution of the AUC values across studies and dosage regimens. The x-axis denotes the dose and study number, for example, 75 mg in study 2 is denoted as 75(s2).

for $i = 1, \ldots, n$, where y_i and x_i are AUC and dose for ith subject, respectively, and I_i is the indicator function such that

$$I_i = \begin{cases} 1 & \text{patient} \\ 0 & \text{healthy volunteer} \end{cases}$$

Figure 47.9 shows regression diagnostic plots indicating that the model adequately characterized the data. The first row panels show a random scatter in the residuals, indicating model adequacy. This is corroborated with the linear relationship between log-transformed AUC and dose (Figure 47.9, bottom left). In the bottom right panel of the figure is a Q-Q plot showing the relationship between the quantiles of residuals and the quantiles of a standard normal distribution indicating the adequacy of the linear model fit. It further strengthens the appropriateness of the assumption of a homoscedastic error structure. Therefore, the regression model with the indicator function was found to be an adequate and appropriate model (13).

The distributions of observed and predicted AUC values are shown in Figure 47.10. The figure shows data from all studies with the inclusion of targeted doses for which knowledge about exposure has been generated.

POOLING OF SAFETY DATA ACROSS STUDIES AND OVERALL EXPOSURE–RESPONSE RELATIONSHIP   The healthy volunteer and diseased subject studies were of different durations. It was necessary to determine if the safety data from individual studies could be pooled for the purpose of integrating the findings. Figure 47.11 shows the clustering together of the results from the different studies, indicating that the pooling of the studies for an overall PM knowledge discovery and creation is valid. The length of treatment duration did not increase the probability of AE occurrence.
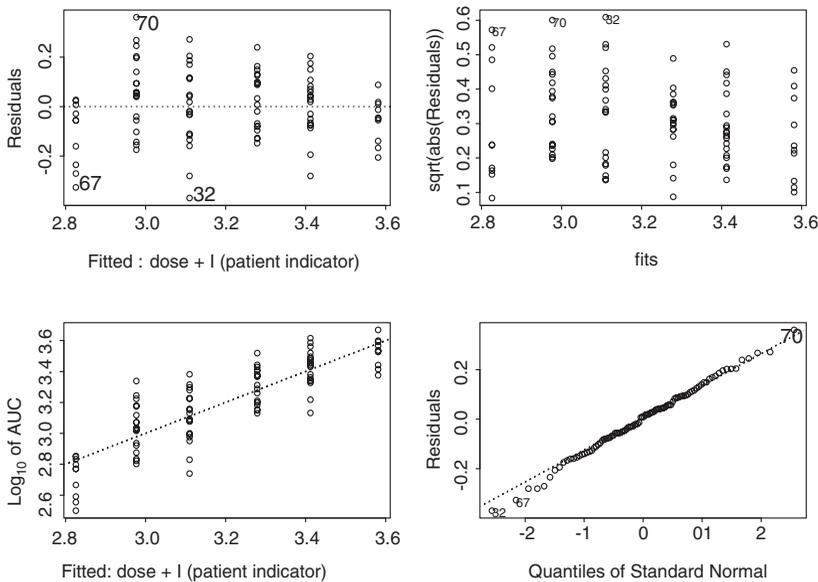
FIGURE 47.9 Diagnostic plots used in the model evaluation of the linear regression model that related log(AUC) to log(dose) incorporating subject type (1 for diseased subjects (patient), and 0 for healthy subject) as covariate. Top row: The left and right panels are residuals showing that the adequacy of the model fit. Bottom row: The left panel plot reinforces the fact that the model adequately describes the data. The right-hand plot shows the adequacy of the error model.
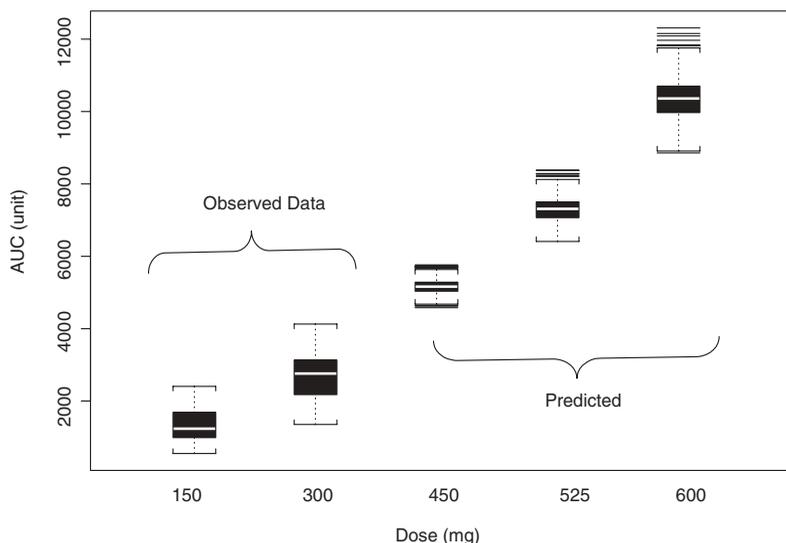
FIGURE 47.10 Distributions of observed and predicted AUC based on the pooled pharmacokinetic data set.
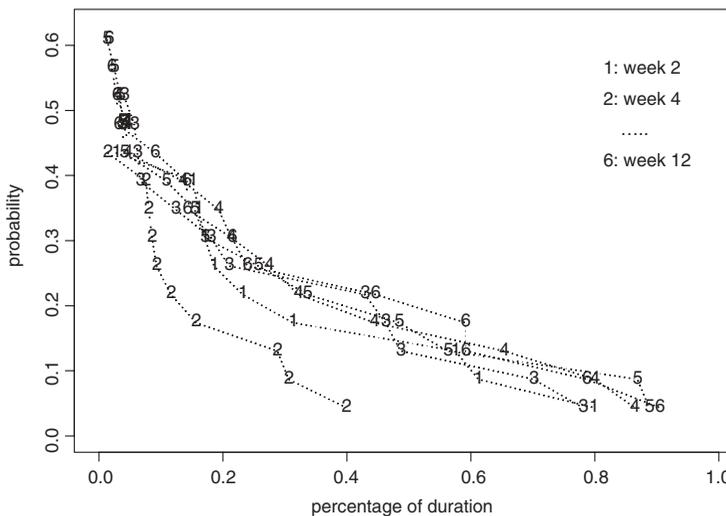
FIGURE 47.11 A biweekly probability plot showing the effect of percentage of duration on the probability of having grade 1 AE. Clustering of the probability plots across studies indicates that it was valid to pool observations across studies for the performance of pharmacometric knowledge discovery and creation.

When the data were pooled, the probabilities for the different durations a subject with a given exposure might spend in grade 1 AE is significantly reduced. For instance, the probability of any occurrence dropped from 100% to 90% when Figures 47.7 and 47.12 are compared—an impact of sample size giving robustness to the outcome. Because of the curvilinear relationship between the explanatory
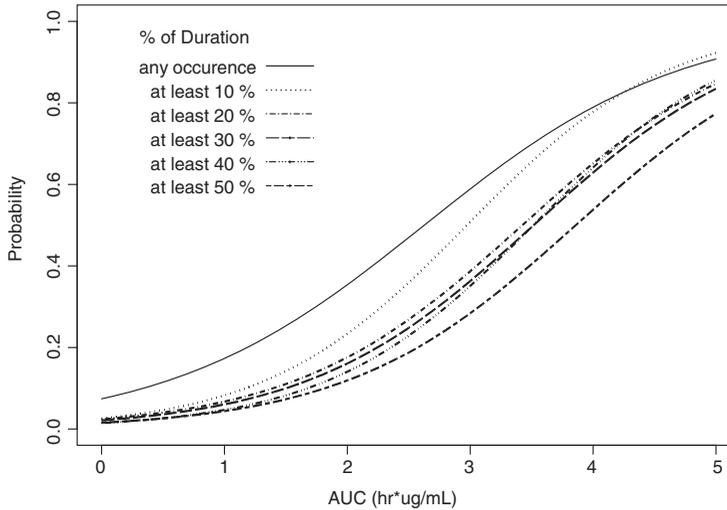
FIGURE 47.12 Exposure–response relationship for the high baseline group showing the probability of having grade 1 AE given different duration percentages. Data were pooled from all three studies comprising healthy and diseased subjects.

variable x and the probability $\pi(x)$, the following population logistic regression model was developed from the pooled data. For a binary response (e.g., reaching grade 1 safety biomarker denoted by yes = 1 or no = 0), the probability $\pi(x)$ is defined as (14)

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \tag{47.2}$$

For this model the odds of having a grade 1 AE is

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) \tag{47.3}$$

This formula provides a basic interpretation for β. The odds increase multiplicatively by e^{β} for every unit increase in x, where x and β can be vectors that represent multidimensional covariates and associated coefficients, respectively. In applying the model to the data, the predictor, x, is AUC and $\pi(x)$ is the probability of having a predefined event (e.g., grade 1 AE occurring with at least 10% of treatment duration).

47.4.3.2 Pharmacometric Knowledge Creation
Having characterized the relationship between the target doses and AUC on one hand, and discovered knowledge about the relationship between AE and AUC from the pooled data across studies on the other hand, the stage was set for creating knowledge about AE that could have been observed with the target doses for a study that had the same duration as study 3. It was assumed that subjects similar to those studied in study 3 would have been exposed to the target doses. This is an

TABLE 47.1 Summary of the Predicted Probabilities Associated with the Occurrence of Different Grades of AE Characterized by the Duration a Subject Spends in a Particular AE Grade[a]

	Mean (SD)			
	450 mg	525 mg	600 mg	300 mg
Probability of Having Grade 1: High Baseline				
Any occurrence	74 (5.03)	76 (4.94)	79 (4.80)	62 (5.76)
10% of Duration	72 (5.74)	75 (5.59)	79 (5.37)	55 (5.37)
20% of Duration	58 (7.35)	62 (7.41)	66 (7.36)	42 (5.3)
30% of Duration	56 (6.25)	60 (6.47)	63 (6.61)	39 (5.11)
40% of Duration	56 (6.98)	60 (7.20)	65 (7.30)	39 (5.22)
50% of Duration	48 (7.83)	52 (8.30)	56 (8.65)	31 (4.57)
Probability of Having Grade 2: High Baseline				
Any occurrence	58 (7.58)	63 (7.71)	67 (7.71)	40 (4.65)
10% of Duration	50 (6.37)	55 (6.77)	59 (7.07)	35 (4.81)
20% of Duration	35 (6.42)	39 (7.04)	42 (7.65)	24 (3.8)
30% of Duration	35 (5.67)	38 (6.28)	41 (6.90)	23 (4.12)
40% of Duration	36 (5.96)	40 (6.56)	44 (7.11)	
50% of Duration	35 (6.47)	38 (7.24)	42 (8.01)	
Probability of Having Grade 3: High Baseline				
Any occurrence	26 (6.87)	29 (7.98)	33 (9.06)	17 (3.53)
10% of Duration	14 (4.08)	15 (4.6)	17 (5.17)	9 (2.69)
20% of Duration	11 (4.43)	12 (5.03)	13 (5.73)	7 (2.62)
30% of Duration				

[a] The 300 mg column observed study data is included as a comparison.

important and strong data supplementation assumption. The assumption enabled the application of the population logistic model developed in the previous section and the generation of exposures for the target doses in Section 47.4.3.1 to obtain the probability of having an AE given the target doses. Table 47.1 provides a summary of the probabilities associated with the occurrence of different grades of AE as characterized by the duration a subject spends in a particular AE grade. As would be expected from the relationship between exposure and AE, a greater severity of AE was predicted for target doses.

47.4.4 The Confluence of Pharmacometric Knowledge Discovery and Creation

A combination of PM knowledge discovery and creation approaches was used to create knowledge about target doses that were not investigated in a previously concluded study. In doing this, an aspect of the response surface in terms of drug safety was characterized. This knowledge was subsequently communicated to provide the mission critical decision making. This application demonstrates a confluence of PM

knowledge discovery and creation approaches in characterizing an aspect of an unexplored region of the response surface.

47.5 SUMMARY

A confluence is the merger or meeting of two or more objects (or subject matters) that seem to inseparably bind their respective forces or attributes into a point of junction. The point of junction of pharmacometric knowledge discovery and creation is in attempting to gain knowledge and understanding of the response surface, especially an unexplored region. Thus, the importance of using a combination of PM knowledge discovery and creation approaches in characterizing an aspect of the response surface in terms of drug safety is discussed.

An overview of some techniques for pharmacometric knowledge discovery is presented together with the introduction of a new metric, the percentile division approach, for data structure revelation; and a new metric (the percentage duration of an AE) for analyzing drug-induced AE is proposed. The application example demonstrates the confluence of PM knowledge discovery and creation in drug development.

REFERENCES

1. M. I. Kleinberg and L. A. Wanke, New approaches and technologies in drug design and discovery. *Am J Health Syst Pharm* **52**:1323–1336; 1341–1343 (1995).

2. C. Connolly, Price Tag for a New Drug: $802 Million; Findings of Tufts University Study Are Disputed by Several Watchdog Groups. *Washington Post*, 1 December 2001.

3. J. A. DiMasi, Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **69**:297–307 (2001).

4. E. I. Ette, P. J. Williams, H. Sun, E. O. Fadiran, F. O. Ajayi, and L. C. Onyiah, The process of knowledge discovery from large pharmacokinetic data sets. *J Clin Pharmacol* **41**:25–34 (2001).

5. P. J. Williams, C. J. Godfrey, A. Roy, H.-M. Chu, and E. I. Ette, Pharmacokinetics/ pharmacodynamics knowledge discovery and creation during drug development, in *Industrial Pharmacokinetics*, P. Bonate and D. Howard (Eds.). AAPS Press, Alexandria, VA, 2004, pp. 163–202.

6. E. I. Ette, H.-M. Chu, and C. J. Godfrey, Data supplementation: a pharmacokinetic/ pharmacodynamic knowledge creation approach for characterizing an unexplored region of the response surface. *Pharm Res* **22**:523–531 (2004).

7. H. Derendorf, L. J. Lesko, P. Chaikin, W. A. Colburn, P. Lee, R. Miller, R. Powell, G. Rhodes, D. Stanski, and J. Venitz, Pharmacokinetic/pharmacodynamic modeling in drug research and development. *J Clin Pharmacol* **40**:1399–1418 (2000).

8. T. Hastie and R. Tibshirani, *Generalized Additive Additive Models*. Chapman B Hall, London, 1990.

9. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, New York, 2001.

10. L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, 1984.

11. E. I. Ette and T. M. Ludden, Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res* **12**:1845–1855 (1995).

12. B. Efron, Nonparametric standard errors and confidence intervals (with discussion). *Can J Stat* **9**:139–172 (1981).

13. E. I. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M.-J. Liu, and E. Capparelli, Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* **43**:610–623 (2003).

14. A. Agresti, *Categorical Data Analysis*. Wiley, New York, 1990, pp. 79–91.

APPENDIX 47.1   CODE FOR PERCENTILE DIVISION APPROACH

```
####################################################
### Percentile Division
####################################################


myQuantile_quantile(data$baseline,prob=seq(0.1,0.7,0.1))

for(i in 1:length(myQuantile))
{
  tmp$baseI_rep(0,nrow(tmp))
  tmp$baseI[tmp$baseline>myQuantile[i]]_1
  cat("\n","****** The baseline Cutoff is",myQuantile[i],"(",i*10,"%)","\n")
  fit.glm4_glm(y ~ x+baseI, family = binomial, data = tmp)
  #print(summary(fit.glm4))
  print(anova(fit.glm1,fit.glm4,test="Chisq"))
}
```

APPENDIX 47.2   CODE FOR THE NEW METRIC CALCULATION

```
##################################################################
### compile new metric.ssc
### to calculat time (days) above any grades (Grade 1 is illustrated here)
### note: AboveTime2.hmc is a customized Splus function
### data set is a.SF.HV
##################################################################


AboveTime2.hmc_function(time, DV, grade)
{
  xy <- data.frame(t = time, y = DV)
  xy <- rbind(xy, xy)
  xy <- xy[sort.list(xy$t), ]
  xy$greater <- rep(0, nrow(xy))
  xy$greater[xy$y >= grade] <- 1
  xy$change <- xy$greater
  xy$newT <- xy$t
  for(j in 2:nrow(xy)) {
```

```
    qq <- xy$greater[j] - xy$greater[j - 1]
    if(qq != 0) {
      c <- (xy$y[j] - xy$y[j - 1])/(xy$t[j] - xy$t[j - 1])
      #slope
      # cat("\n"," slope",j,"..",c,"\n")
      if(qq == 1) {
        xy$newT[j - 1] <- xy$t[j - 1] + (grade - xy$
         y[j - 1])/c
        xy$change[j - 1] <- 1
      }
      if(qq == -1) {
        xy$newT[j] <- xy$t[j - 1] + (grade - xy$y[
         j - 1])/c
        # cat("\n"," (grade[i]-xy$y[j-1])/c",round((grade[i]-xy$y[j-1])/c,2),
"\n")
        # cat("\n","xy$t[j-1]",xy$t[j-1],"
","xy$newT[j]",round(xy$newT[j],2),"\n")
        xy$change[j] <- 1
      }
    }
  }
  xy <- xy[xy$change == 1, ]
  d <- seq(1, nrow(xy), 2)
  #print(d)
  xy$diff <- rep(NA, nrow(xy))
  d1 <- xy$newT[d + 1]
  d2 <- xy$newT[d]
  d3 <- d1 - d2
  #print(d3)
  xy$diff[d] <- d3
  xy
}


#################################
myGrade_c(25,31.6,52.6,105.1)

### for time above grade 1
ID_unique(SF.combined$ID[SF.combined$SF>myGrade[1]])
print(ID)
ID_ID[!is.na(ID)]
print(ID)
a.SF.HV$t.g1_rep(NA,nrow(a.SF.HV))

for(i in 1:length(ID))
#for(i in 1:1)
{
  if (ID[i]<3000){myDay_7}
    else {myDay_14}
```

```
 b_SF.combined[SF.combined$ID==ID[i] & SF.combined$day>=0 &
 SF.combined$day<=myDay,]
 b_b[sort.list(b$day),]
 if(is.na(b$SF[b$day==0]))
   { b$SF[b$day==0]_a.SF.HV$baseline[a.SF.HV$ID==ID[i]]}
 b_b[!is.na(b$day) & !is.na(b$SF),]
 if( any( b$SF[b$SF>myGrade[1] ] )==T )
 {
   xy_AboveTime2.hmc(time=b$day,DV=b$SF,grade=myGrade[1])
   a.SF.HV$t.g1[a.SF.HV$ID==ID[i]]_sum(xy$diff,na.rm=T)
 }
}
```

APPENDIX 47.3   DATA PROCESSING CODE

```
###################################
# data set compiling and preprocessing
# source data set is named data
###################################
data$study_round(data$ID/1000,0)
data$regimen_rep(NA,nrow(data))
data$r_rep(NA,nrow(data))

data$regimen[data$study==2 & data$Dose==150]_"150 TID"
data$regimen[data$study==2 & data$Dose==450]_"450 BID"
data$regimen[data$study==3 & data$Dose==75]_"75 BID"
data$regimen[data$study==3 & data$Dose==150]_"150 BID"
data$regimen[data$study==3 & data$Dose==300]_"300 BID"
data$regimen[data$Dose==0]_"Placebo"

# to create "r" factor column to indicate the dose regimens across studies
data$r[data$study==2 & data$Dose==150]_3
data$r[data$study==2 & data$Dose==450]_5
data$r[data$study==3 & data$Dose==75]_1
data$r[data$study==3 & data$Dose==150]_2
data$r[data$study==3 & data$Dose==300]_4
data$r[data$Dose==0]_0
data_data[sort.list(data$r),]

## data1 is the transformed data object
## SF is the safety biomarker column
data1_data[data$DAY==7,]
data2_data[data$DAY==1,]
data1$baseline[match(data2$ID,data1$ID)]_data2$SF
data1$day7_data1$SF
data2_data[data$DAY==14,]
data1$day14[match(data2$ID,data1$ID)]_data2$SF
```

```
## to translate baseline safety biomarker values into baseline grades (0-4)
data1$baseline.grade_rep(NA,nrow(data1))
data1$baseline.grade[data1$baseline>=0 & data1$baseline<25 ]_0
data1$baseline.grade[data1$baseline>=25 & data1$baseline<31.5 ]_1
data1$baseline.grade[data1$baseline>=31.5 & data1$baseline<52.5 ]_2
data1$baseline.grade[data1$baseline>=52.5 & data1$baseline<105 ]_3
data1$baseline.grade[data1$baseline>=105 ]_4


## to translate Day 7 safety biomarker values into grades (0-4)
data1$day7.grade_rep(NA,nrow(data1))
data1$day7.grade[data1$day7>=0 & data1$day7<25 ]_0
data1$day7.grade[data1$day7>=25 & data1$day7<31.5 ]_1
data1$day7.grade[data1$day7>=31.5 & data1$day7<52.5 ]_2
data1$day7.grade[data1$day7>=52.5 & data1$day7<105 ]_3
data1$day7.grade[data1$day7>=105 ]_4


## to calculate if day 7 grades change from baseline grades
data1$diff_data1$day7.grade-data1$baseline.grade
data1$response_rep(NA,nrow(data1))
data1$response[data1$diff==0]_"no"
data1$response[data1$diff>0]_"yes"


data1 <- convert.col.type(target = data1, column.spec = list("regimen"),
  column.type = "factor")
```

APPENDIX 47.4   PAIRS PLOT

```
############################################################
# Figure 3 (pairs plot)
############################################################


aa_data[data$dose>0, c("Cmax","AUC","SF_baseline","SF_Day14")]
pairs(aa)
```

APPENDIX 47.5   TREE BASED MODEL

```
############################################################
# Figure 4 (Tree based model)
############################################################
a_data[data$dose>0,c(2:10,13)]
a_na.omit(a)
a$grade_rep(NA,nrow(a))
a$grade[a$Day14>=25]_1
a$grade[a$Day14<25]_0
```

```
b <- tree(grade ~ ., data = a[,-10])
plot(b)
text(b)
print(b)
```

APPENDIX 47.6   2-D BUBBLE PLOT

```
#############################################################
# Figure 5 (SF grade vs Cmax)
#############################################################

data2_data1[ !is.na(data1$diff),]  # data2 contains subjects which have both
  day 1 & 7 SF scores


par(mfrow=c(1,1))
plot(x=data2$Cmax,y=data2$baseline,type="n",xlab="Exposure Parameter",
     ylab="Safety Biomarker Baseline",xlim=c(-20,1750),ylim=c(0,23))


# to plot Cmax based on studies

data3_data2[data2$study==2,]
points(x=data3$Cmax,y=data3$baseline,pch="1",col=3)
data3_data2[data2$study==3,]
points(x=data3$Cmax,y=data3$baseline,pch="2",col=6)

data3_data2[data2$response=="yes" & data2$study==2,]
for(i in 1:nrow(data3))
{
  if (!is.na(data3$Cmax[i]))
  {
    points(x=data3$Cmax[i],y=data3$baseline[i],pch=1,cex=1.5+data3$diff[i])
  }
}

data3_data2[data2$response=="yes" & data2$study==3,]
for(i in 1:nrow(data3))
{
  if (!is.na(data3$Cmax[i]))
  {
    points(x=data3$Cmax[i],y=data3$baseline[i],pch=1,cex=1.5+data3$diff[i])
  }
}

## to plot placebo
data3_data1[!is.na(data1$baseline) & data1$Dose==0,]
```

```
points(x=rep(-20,nrow(data3)),y=data3$baseline,pch="p")
data3_data1[data1$response=="yes" & data1$Dose==0,]
if (nrow(data3)>0)
{
  for(i in 1:nrow(data3))
  {
    points(x=-20,y=data3$baseline[i],pch=1,cex=1.5+data3$diff[i])
  }
}


#title("SF grade changes at day 7")


# to plot division lines
lines(y=c(7.4,25),x=rep(579.5,2),lty=2,lwd=4)
#lines(y=c(0,25),x=rep(1500,2),lty=2,lwd=4)
lines(x=c(-20,1500),y=rep(7.4,2),lty=2,lwd=4)
lines(y=c(0,10),x=rep(1500,2),lty=2,lwd=4)
key(text=list(c("G1","G2","G3"),font=4),points=list(pch=c(1,1,1),cex=c(1:3)
  +1.5),between=4,
  corner=c(1,1),title="Grade",cex.title=1.2,border=1)
```

APPENDIX 47.7   AN EXAMPLE CODE FOR LOGISTIC REGRESSION

```
#####################################################
## Logistic regression using a binary response vector.
##
## subjects that have grade 1 safety biomarker value
## t.g1 column is the time above grade 1 during treatment
## x column is the chosen exposure parameter
#####################################################


tmp_data
tmp$x_tmp$AUC
tmp$y_rep(0,nrow(tmp))
tmp$y[tmp$t.g1>0]_1
tmp_tmp[!is.na(tmp$x),]


fit.glm1_glm(y ~ x, family = binomial, data = tmp)
fit.glm2_glm(y ~ x+baseline, family = binomial, data = tmp)
fit.glm3_glm(y ~ x+baseline+sex, family = binomial, data = tmp)
cat("\n","*****************************","\n")


anova(fit.glm1,fit.glm2,fit.glm3)
summary.aov(fit.glm3)
```

APPENDIX 47.8   AUC PLOT

```
############################################################
### Figure 8 (AUC across studies)
### data set is ExposureAll that contains AUC and dose2 (dosage
### regimens) columns
############################################################

par(mfrow=c(1,1))
#boxplot(split(ExposureAll$AUC, ExposureAll$dose2))
unique(ExposureAll$dose2)
tmp10_ExposureAll
tmp10$x_rep(NA,nrow(tmp10))
x1_c("75(s2)","150(s1)","150(s2)","150(s3)","300(s2)","300(s3)","450(s1)")

for(i in 1:length(x))
{
  tmp10$x[tmp10$dose2==x[i]]_i
}

boxplot(split(log10(tmp10$AUC),tmp10$x),names=x1,axes=F)
y_c(300,500,750,1250,2000,3000,5000)
axis(2,at=log10(y), labels=as.character(y),ticks=T,srt=90)
box()
#mtext(side=2,"AUC in log10 scale",line=3,cex=1.2)
mtext(side=2,"AUC (unit)",line=3,cex=1.2)
mtext(side=1,"Dose mg (Study)",line=3,cex=1.2)
```