A MANAGER'S GUIDE TO

THE DESIGN AND CONDUCT OF CLINICAL TRIALS

PHILLIP I. GOOD

A MANAGER'S GUIDE TO THE DESIGN AND CONDUCT OF CLINICAL TRIALS

Phillip I. Good, Ph.D.

WILEY-LISS

A JOHN WILEY & SONS, INC., PUBLICATION

A MANAGER'S GUIDE TO THE DESIGN AND CONDUCT OF CLINICAL TRIALS

A MANAGER'S GUIDE TO THE DESIGN AND CONDUCT OF CLINICAL TRIALS

Phillip I. Good, Ph.D.

WILEY-LISS

A JOHN WILEY & SONS, INC., PUBLICATION

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

Copyright © 2002 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or mechanical, including uploading, downloading, printing, decompiling, recording or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

This title is also available in print as ISBN 0-471-22615-7.
For more information about Wiley products, visit our web site at www.Wiley.com.

This book has benefited from a number of reviewers. Those I'm at liberty to name are Bernarr Pardo for his excellent advice on data management and David Salsburg.

Contents

PART ONE PLAN 1

1 Cut Costs and Increase Profits **3**
  No Excuse for the Wastage **3**
  Front-Loaded Solution **4**
  Downsizing **5**
    A Final Word **5**

2 Guidelines **9**
  Start with Your Reports **9**
  The Wrong Way **11**
  Computer-Assisted Data Entry **11**
  Don’t Push the River **12**
  KISS **12**
  Plug the Holes as They Arise **14**
  Pay for Results, Not Intentions **14**
  Plan, Do, Then Check **14**

3 Prescription for Success **15**
  Plan **15**
  Do **16**
  Check **18**

4 Staffing for Success **19**
  Design Team **19**
  Regulatory Approval **21**
  Implementation Team **21**

- Data Entry Software **22**
  - Test the Software **22**
- Investigator Panels **23**
- Site Coordinators **24**
- External Review Committees **24**
  - Recruit and Enroll Patients **25**
  - Conduct the Trials **25**
- Programs for Data Analysis **26**
- The People You Don’t Need **28**

5 Design Decisions **29**

- Should the Study Be Performed? **30**
- Study Objectives **30**
- Primary End Points **32**
  - Secondary End Points **34**
- Baseline Data **35**
- Who Will Collect the Data **36**
- Quality Control **36**
- Study Population **37**
- Timing **39**
- Closure **40**
  - Planned Closure **40**
  - Unplanned Closure **41**
- Be Defensive. Review. Rewrite. Review Again **42**
- Checklist for Design **43**
- Budgets and Expenditures **44**
- For Further Information **44**

6 Trial Design **47**

- Baseline Measurements **48**
- Controlled Randomized Clinical Trials **48**
  - Randomized Trials **50**
  - Blocked Randomization **50**
  - Stratified Randomization **51**
- Single- and Double-Blind Studies **52**
  - Exceptions to the Rule **54**
- Sample Size **54**
  - Which Formula? **54**
  - Precision of Estimates **55**
- Number of Treatment Sites **60**
- Alternate Designs **60**

Taking Cost into Consideration **62**
For Further Information **63**

7 Exception Handling **65**
  Patient Related **65**
    Missed Doses **65**
    Missed Appointments **65**
    Noncompliance **66**
    Adverse Reactions **66**
    Reporting Adverse Events **66**
    When Do You Crack the Code? **67**
  Investigator Related **68**
    Lagging Recruitment **68**
    Protocol Deviations **68**
    Site-Specific Problems **69**
    Closure **70**
    Intent to Treat **70**
  Is Your Planning Complete? A Checklist **71**

PART TWO DO 73

8 Documentation **75**
  Guidelines **75**
  Initial Submission to the Regulatory Agency **76**
    Sponsor Data **76**
    Justifying the Study **76**
    Objectives **77**
    Patient Selection **77**
  Treatment Plan **78**
  Outcome Measures and Evaluation **79**
  Procedures **79**
    Clinical Follow-Up **79**
  Adverse Events **80**
  Data Management, Monitoring, Quality Control **80**
  Statistical Analysis **80**
  Investigator Responsibilities **81**
  Ethical and Regulatory Considerations **82**
  Study Committees **82**
  Appendixes **82**
  Sample Informed Consent Form **83**
  Procedures Manuals **84**

    - Physician's Procedure Manual **85**
    - Laboratory Guidelines **86**
  - Interim Reports **86**
    - Enrollment Report **86**
    - Data in Hand **87**
    - Adverse Event Report **87**
    - Annotated Abstract **88**
  - Final Report(s) **91**
    - Regulatory Agency Submissions **91**
    - E-submission **92**
    - Journal Articles **93**
  - To Learn More **94**

9 Recruiting and Retaining Physicians and Patients **95**
  - Recruiting Physicians **95**
    - Teaching Hospitals **96**
    - Clinical Resource Centers **97**
    - Look to Motivations **97**
    - Physician Retention **98**
    - Get the Trials in Motion **98**
  - Patient Recruitment **99**
    - Factors in Recruitment **99**
    - Importance of Planning **100**
    - Ethical Considerations **101**
    - Mass Recruiting **101**
  - Patient Retention **102**
    - Ongoing Efforts **103**
    - Run-in Period **104**
  - Budgets and Expenditures **105**
  - To Learn More **105**

10 Computer-Assisted Data Entry **109**
  - Pre-data Screen Development Checklist **110**
    - Develop the Data Entry Software **110**
    - Avoid Pre-defined Groupings **112**
  - CDISC Submission Standards **112**
  - Screen Development **114**
    - Radio Button **114**
    - Pull-down Menus **116**
    - Type and Verify **116**
    - When the Entries Are Completed **117**
    - Audit Trail **117**

- Electronic Data Capture **118**
- Testing **119**
  - Formal Testing **120**
  - Stress Testing **121**
- Training **122**
  - Support **123**
- Budgets and Expenditures **124**
- To Learn More **124**

11 Data Management **125**
- Options **125**
  - Flat Files **125**
  - Hierarchical Databases **127**
  - Network Database Model **128**
  - Relational Database Model **128**
  - Which Database Model **131**
  - Object-Oriented Databases **132**
- Clients and Servers **132**
  - One Size Does Not Fit All **133**
- Combining Multiple Databases **133**
  - The Key Is the Key **134**
  - Transferring Data **136**
- Quality Assurance and Security **137**
  - Maintaining Patient Confidentiality **137**
  - Access to Files **138**
  - Maintaining an Audit Trail **139**
  - Security **139**
- For More Information **140**

12 Are You Ready? **141**
- Pharmaceuticals/Devices **141**
- Software **142**
- Hardware **142**
- Documentation **142**
- Investigators **142**
- Review Committees **143**
- Patients **143**
- Regulatory Agency **143**
- Test Phase **143**

13 Monitoring the Trials **145**
- Roles of the Monitors **145**

- Before the Trials Begin **147**
- Kick-off Meetings **148**
- Duties During Trial **148**
  - Site Visits **149**
  - Between Visits **150**
  - Other Duties **152**
- Maintaining Physician Interest in Lengthy Trials **152**

14 Managing the Trials **155**
- Recruitment **156**
  - Late and Incomplete Forms **157**
  - Dropouts and Withdrawals **157**
  - Protocol Violations **158**
  - Adverse Events **158**
- Quality Control **159**
- Roles of the Committees **160**
- Termination and Extension **161**
- Extending the Trials **162**
- Budgets and Expenditures **163**
- To Learn More **164**

15 Data Analysis **165**
- Report Coverage **165**
- Understanding Data **166**
  - Categories **166**
  - Metric Data **168**
- Statistical Analysis **170**
  - Categorical Data **172**
  - Ordinal Data **173**
  - Metric Data **174**
  - Time-to-Event Data **176**
- Step by Step **179**
  - The Study Population **179**
  - Reporting Primary End Points **180**
  - Exceptions **181**
  - Adverse Events **183**
- Analytical Alternatives **183**
  - When Statisticians Can’t Agree **184**
  - Testing for Equivalence **185**
  - Simpson’s Paradox **186**
  - Estimating Precision **187**
- Bad Statistics **189**

        - Using the Wrong Method **189**
        - Choosing the Most Favorable Statistic **189**
        - Making Repeated Tests on the Same Data **190**
        - Ad hoc, Post hoc Hypotheses **191**
    - Interpretation **193**
    - Software Documentation **193**
    - To Learn More **195**
    - A Practical Guide to Statistical Terminology **197**

PART THREE CHECK 199

16 Check **201**
- Closure **201**
    - Patient Care **201**
    - Data **202**
    - Spreading the News **202**
- Postmarket Surveillance **202**
- Budget **202**
    - Variable and Fixed Expenditures **203**
- Controlling Expenditures **203**
- Trial Review Committee **203**
- After Action Review **204**
    - Interactions **206**
    - Adverse Events **206**
- Collateral Studies **207**
- Future Studies **207**
    - Data **207**
    - Patients **208**
- To Learn More **208**

Appendix Software **209**
- Choices **209**
    - All in One **209**
    - Project Management **210**
    - Almost All in One **210**
- Data Entry **211**
    - Handheld Devices **211**
    - Touch Screen **211**
    - Speech Recognition **211**
    - e-CRFs **211**
    - Do It Yourself **211**

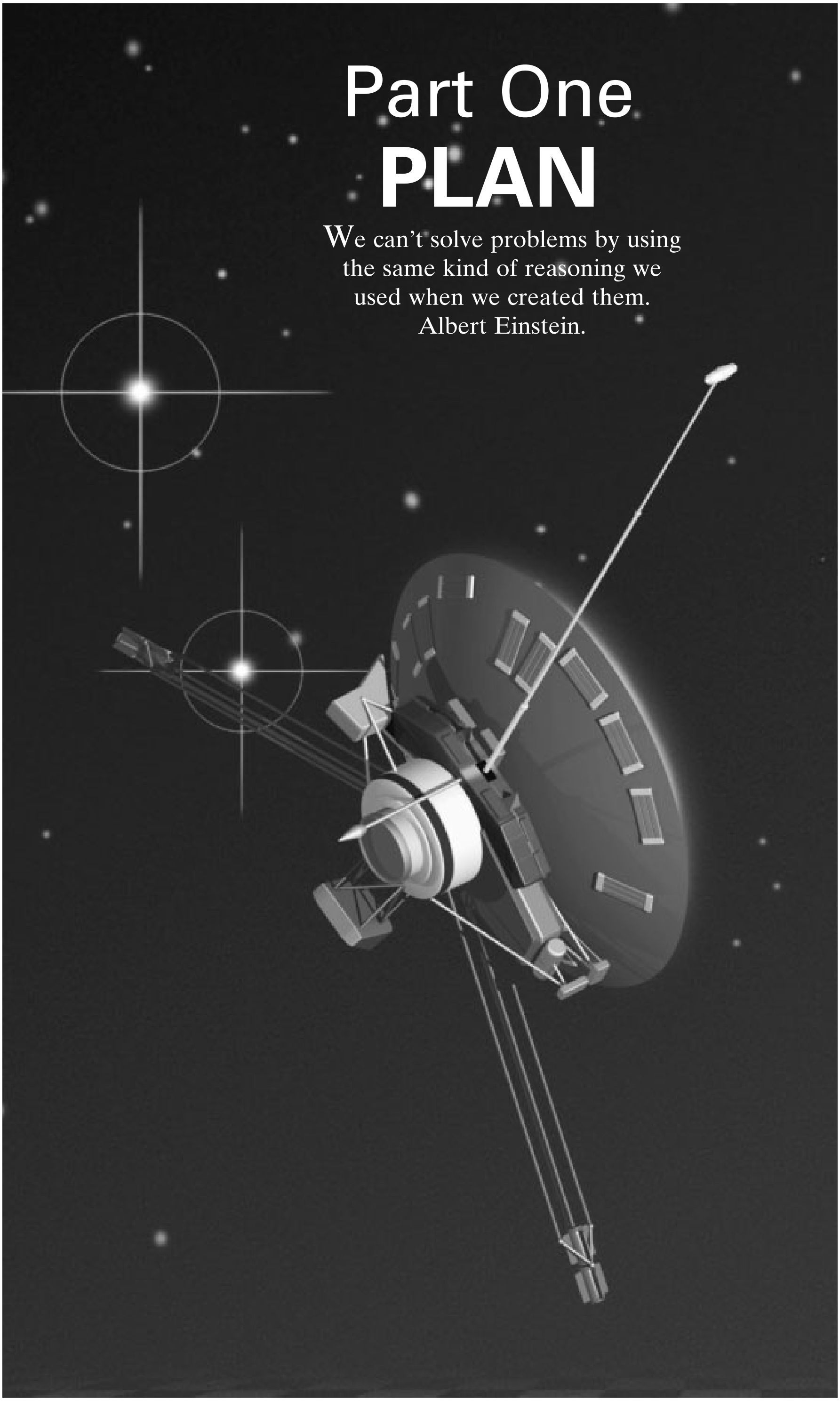
    - Via Web Access **212**
  - Data Management **212**
  - Date Entry and Data Management **213**
    - Small-Scale Clinical Studies **213**
    - Clinical Database Managers **213**
  - Data Analysis **214**
  - Utilities **215**
    - For Sample Size Determination **215**
    - Screen-Capture **215**
    - Data Conversion **216**

Author Index 217
Subject Index 221

Part One

PLAN

We can't solve problems by using
the same kind of reasoning we
used when we created them.
Albert Einstein.



Chapter 1

Cut Costs and Increase Profits

THE ESSENCE OF THE GUIDELINES presented here—start with your reports, enter the data directly into the computer, validate on entry, and monitor your results continuously—first appeared in a newsletter I edited in the mid 1980s. The reactions of readers then ranged from tepid to outwardly hostile: "We can't afford to give every physician a computer," raged one data manager, ignoring the $10,000 per patient that is the normal minimal expense for clinical data. "What will become of all the people we've trained as encoders?" bemoaned another months before the furious downsizing that characterized the late 1980s.

Such reactions make even less sense today when desktop computers are available for less than a $1000 apiece (and even lower priced when purchased 25 or a 100 at a time) and every corporation is leaner and meaner than it has ever been. Yet everywhere we look the same old-fashioned outmoded and hopelessly inefficient procedures are still in place.

NO EXCUSE FOR THE WASTAGE

There is no excuse for wastage and only one explanation: middle management in pharmaceutical and device companies has focused on their own survival, not the corporation's. They have minimized risks by doing what was done before and have placed the company at risk in consequence. They have developed elaborate time-consuming schemes to make today's paperless system function as if we still had to carve out each letter by hand and cost their companies millions in unnecessary added costs and millions more in lost profits because of the delays.

And why the delays? So our manager won't rock the boat, be caught innovating, or, worse, bring on board persons with skills that fail to match existing job descriptions.

But the bottom line is that electronic data capture coupled with careful monitoring will cut costs and shorten the time to realizing profit.

FRONT-LOADED SOLUTION

This text is about a great deal more than computer-aided data entry. The essence of the solution is that we need to spend far more time on planning, less on the repairs.

My pessimism stems, in part, from my having spent the last 20 years as a consultant to drug and device firms. As a consultant, I was always called in at the last moment to "fix" the problem. The "fix" took months, was generally unsatisfactory, and all hope of profit vanished when the competitor was first to market.

I worked full time once, too, for a fast-track boss who'd earned his spurs as a firefighter. He put down every preventive measure I proposed. But then, what's a firefighter without a fire?

The solutions offered to you here are front-loaded and may seem expensive. But by putting in the preventive planning effort now, your company will avoid far more time-consuming and expensive delays later.

Anyone who has ever spent much time on the water (or in the air) knows that once underway it is far better to under- than to oversteer. On the other hand, no experienced sailor (or aviator) would consider getting underway without first making sure all systems were fully functional and life jackets, life raft, and emergency rations ready if needed.

I can understand and occasionally sympathize when biotech startups attempt to cut corners by doing the absolute minimum until they (and, more important, their investors) can be confident the project will be successful. It ends up costing these companies and their investors more in the end—not infrequently, an entire set of trials must be repeated all over from the beginning—but if you don't have money to begin with and must

TWO APPROACHES TO MANAGEMENT

1. Tentative but responsible, avoid precipitous action and waits to see how the situation will develop before intervening.
2. Envisions the worst and plans for it.

Effective managers employ both.

wait upon the necessary venture capital, what choice do you have?

The puzzle comes when a large well-capitalized firm makes the same errors, errors that can only be attributed to poor management and slothful minds that simply hope to defer the inevitable.

DOWNSIZING

Take downsizing as one example of sloppy management. Too often downsizing has taken place by percentages and not in terms of the skills the modern corporation needs. Stand back, see what you are trying to accomplish, then hire, or, better still, retrain in accordance with current requirements.

Developing all the details of safety and efficacy assessment, data gathering, and recruitment before one begins demands time and patience. The counterargument that one cannot foresee every contingency is largely false. The fact is that when one is forced to lay out all the elements of a design before commencing a study, not infrequently one manages to foresee 99.9% of the potential problems. Throwing up your hands and crying, "it's just too difficult, let's wait for the data," is the act of a child, not of a mature manager.

Often, those in upper management cannot understand the delay. Yet the tale of the ever befuddled pharmaceutical and device company told in the chapters that follow is too often the case in all too many clinical studies. The high price of pharmaceuticals today masks the costs of ineptitude.

Electronic data capture cuts costs and shortens the time to realizing profit.

A Final Word

For the vast majority of readers, no explanation of why we do clinical trials is necessary. Supervising or participating in clinical trials may even be your primary occupation. But there are a few of you, inventors and entrepreneurs, who are asking just why your drug/device can't be marketed without expensive trials. It's been tested in the lab: you know it works.

The obvious reason is that the regulatory agency won't let you market your intervention without them. But there is a greater, more important motivation: without an organized well-controlled randomized clinical trial, a single run of bad luck, a whim of fate, could forever deny the public of a promising cure and you and your company of justified profits. Think of the controversy surrounding silicon implants. Women got sick, sued, and won millions in damages without the slightest scientific evidence supporting their claims. Manufacturers went bankrupt; hundreds of women had (as it proved, unnecessary) surgery to remove the implants. Yet these bankruptcies could have been avoided had the manufacturers of that period sponsored a well-controlled clinical trial.[1]

For your product to achieve the full success it deserves, you need to know what kind of individuals will respond best to the new treatment and what kind would do best to avoid it. Controlled large-scale clinical trials are the *only* way to get the answers you need.

IN THE NEWS

"German prosecutor launches probe against Bayer over how the company handled the withdrawal of Baycol/Lipobay, the anti-cholesterol drug that has been linked to fatal side effects."
Financial Times, 4 September 2001. (In the litigious United States, just call 1-877-Toxic-RX to become part of a class action suit against Bayer.)

"Baxter recalls blood filters after deaths."
***Financial Times*, 4 September 2001.**

"Class action suit challenges Pfizer over the way it conducted clinical trials in Nigeria five years ago."
***Financial Times*, 3 September 2001.**

Aspirin is unparalleled for its ability to ease pain, lower fever, and suppress inflammations. I carry a couple of aspirin with me in the car because I've read that taking an aspirin during or just after a heart attack could save my life. But if I were already taking an anticoagulant, an aspirin could mean death.

On the back of the aspirin bottle, in large bold print, much larger than the other writing you'll find on the label, are the words, "It is especially important not to use aspirin during the last three months of pregnancy unless specifically directed to do so by a doctor because it may cause problems in the unborn child or complications during delivery." Important words that when written in the language of the potential consumer will forestall lawsuits.[2]

In what follows, we provide guidelines for your trials and a prescription for success. We tell you the contingencies you need to plan for and the design decisions you need to make. We show you how to conduct and monitor long-term clinical trials and, finally, how to review the results so that you can be still more effective in the trials of your next successful product.

Every profession likes to cloak their actions, even the simplest, in arcane language virtually unintelligible to outsiders (statisticians and

---

[1] Angell, M. (1996), *Science on Trial*: *The Clash of Medical Evidence and the Law*, New York: Norton.

[2] *Ramirez v. Plough, Inc.*, 6 Cal.4th 539.

CLINICAL TRIALS

Clinical trials consist of a randomized comparison over a fixed period of time of an intervention method (drug, device, or biological alteration) of interest against an established standard or a negative control (placebo).

The trials are normally preceded by in numero (computer), in vitro (cell culture), and in vivo (animal) experiments, both acute (one time) and chronic (over an extended period), and, in some cases, retrospective studies of the effects of the intervention in humans.

The initial Phase I or *safety trials* focus on the potential adverse effects of the intervention in humans. In the case of drugs and biologics, these trials are used to establish maximum acceptable dose levels (the minimum toxic dose). They generally involve only a small number of subjects and a one-time or short-term intervention. An extended period of several months may be used for follow-up purposes.

The subsequent Phase II or *efficacy trials* are used to establish minimum effective dose levels and to obtain some idea of the nature of secondary responses to the intervention and possible adverse side effects.

The focus of this text is the final or Phase III clinical trials. These involve large numbers of subjects (500 to 5000), studied over an extended period of time (two to five years*) with the possibility of an even longer ongoing follow-up. The larger number of subjects in this type of trial provides an opportunity to study the effects of the intervention on different subgroups (women as well as men, smokers as well as nonsmokers, diabetics and nondiabetics) and to assess the effects of concurrent medications and various risk factors on the ultimate outcome. The longer time period provides for an assessment of the effects of chronic usage along with any other long-term effects.

*Phase III trials devoted soley to efficacy can be considerably shorter.

computer scientists are particular offenders). We've tried our best to describe the work of the innumerable specialists in terms all can understand. My articles have appeared in airline magazines, *Sports Now*, *Volleyball Monthly*, and a half-dozen newspapers. Hopefully, you'll understand everything written, the first time through.

I'd recommend that you read this book twice though: the first time to get an overview, and the second (and, perhaps, the third) time on a chapter-by-chapter basis as each stage in your trials arises. Each chapter contains checklists, so you might want to retain a copy of this book for yourself and put a second copy in the hands of the specialist who will be carrying out that chapter's functions.

Specialists (even statisticians and computer programmers) will also find this text of interest, not only for the checklists and lists of further readings that come with each chapter but because this book covers and, hopefully, clarifies the activities of all the other members of the project team.

Thanks for reading.

Phillip Good, PhD
Huntington Beach, CA USA
pigood@oco.net

Chapter 2

Guidelines

THE PURPOSE OF THIS CHAPTER AND THIS TEXT is to provide you the manager with a set of guidelines for the successful design and conduct of clinical trials:

- **Start with your reports**
- **Use computer-assisted data entry**
- **Don't push the river**
- **KISS**
- **Plug the holes as they arise**
- **Pay for results, not intentions**
- **Plan, do, check**

START WITH YOUR REPORTS

Let your objectives determine the data you will collect. Too often, data collection forms arise as the result of brain storming by a committee. Ten questions on three forms for John's group, and so forth. (See sidebar.) The correct, effective way for study design is to list each of the study objectives, then backtrack to the data needed to perform the necessary calculations.

With price tags well into the millions, clinical studies today are not an academic exercise. The data to be collected should be determined by the objectives of the study and not the other way around.

Begin by printing out a copy of the final report(s) and the package insert you would like to see:

> **743 patients self-administered our psyllium preparation twice a day over a three-month period. Changes in the Klozner-Murphy self-satisfaction scale over the course of treatment were compared with**

> **those of 722 patients who self-administered an equally foul-tasting but harmless preparation over the same time period.**
>
> **All patients in the study reported an increase in self-satisfaction, but the scores of those taking our preparation increased an average of 2.3 + 10.5 points more than those in the control group.**
>
> **Adverse effects included . . .**

These reports will determine the data you need to collect. Your list of potential adverse effects should be based on a review of past studies with your drug/device and with other agents in the same class of drug/device. In some instances, for example, when you want to demonstrate that your treatment is as efficacious as the standard but has fewer, less severe side effects, adverse effects should be a second or even a primary focus of your study.

Do not hesitate to write in exact numerical values for the anticipated outcomes, your best guesses. These guesstimates, for efficacy and for adverse effects, will be needed when determining sample size. (See Chapter 6.) Make sure you've included all end points and all

COLLECT ONLY THE DATA YOU NEED

"We need to cut costs." I was told by a group concerned with nicotine addiction. Could I help them develop a budget for a forth-coming feasibility study? They had ambitious plans. "We're going to collect the following data for each patient:

- Kansas Family Satisfaction Scale (4 questions)
- Motivation for Counseling Scale (24 questions)
- Internal Control Index (13 questions)
- Penn State Worry Questionnaire (14 questions)
- Nicotine Withdrawal Symptoms Scale (16 questions)
- Smoking and prior cessation history.

  Vital signs including
- Blood pressure, pulse, temperature, and respiration.
- Laboratory results (electrolytes, hematology, creatinine, BUN, glucose fasting, pregnancy test results, and CO breath results)
- EKG findings."

I told them I was very impressed—we consultants lie a lot—then asked what they expected to include in their final report?

"The number of subjects that stopped smoking or reduced their smoking by 50%. And their withdrawal symptoms after six months."

"Will you be checking nicotine levels by a urinalysis?"

"Can't afford it."

"You can't afford what you don't need either. And you don't need to collect data on anything other than a smoking and prior cessation history and the nicotine withdrawal symptoms scale."

We had our budget.

anticipated side effects in your hypothetical report. Once this prototype report is fleshed out, you'll know what data you need to collect and will not waste your company's time on unnecessary or redundant effort.

THE WRONG WAY

The wrong way to plan a study is to begin with the forms that were used in a previous set of trials. The sole reason for such a choice, regardless of all proffered rationalizations, is to shift the blame in case something goes wrong. The forms from a previous study seldom make even a good "starting point" (an often heard suggestion). Would you line up for the 100-meter hurdles where you stood to hurl the javelin? Never mind where the starting point used to be, locate the finish line, then back up 100 meters.

When and if you need to ask some of the same questions asked in a previous study, take advantage of your past experience to prepare questions that are unambiguous with definitions that exhaust all the possibilities.

COMPUTER-ASSISTED DATA ENTRY

Computer-assisted data entry, that is, electronic data capture at the physician's workplace (hospital or office) and at the time of the patient's visit or procedure, offers at least four advantages over paper forms:

1. **Immediate detection and correction of errors**
2. **Reduced sources of error**
3. **Open-ended, readily modified forms**
4. **Early detection of trends and out-of-compliance sites**

Data entry can be via keyboard, touch-screen, voice-recognition, or a hand-held device. Of course, the mandatory paper form required by many regulatory agencies is easily printed for signature afterward.[3]

It is well known that the ability to correct bad data declines exponentially with the time elapsed between the observation and the correction. By providing for validation of data at the time of entry, typographical and other errors can be trapped and corrected immediately.

---

[3] In the United States, 21 CFR, part 11, provides for the use of electronic signatures.

Abnormal values (e.g., a blood pressure of 80 over 40) would require confirmation at the time of entry or could be rejected altogether (e.g., a BP of 12 over 8). Only subtle typographical errors would slip through—an 85 instead of an 86, but never an 85 instead of an 8.5. Missing values are forestalled.

Computer-assisted data entry means there are no intermediate forms—nothing is scrawled on the backs of envelopes or left to memory—nor are there errors in transcription, in copying and recopying.

Computer-assisted data entry facilitates monitoring and allows you to stay on top of problems. Once the data are in the computer, they can be collated back at your facility with data from other patients and examined for trends. If you detect ambiguities in questions or an overused "other" category during the tryout phase, the electronic form is easily modified. You can also determine which sites, if any, may require additional assistance.

DON'T PUSH THE RIVER

More time is wasted by pharmaceutical and medical device firms in trying to fight, circumvent, or outwit government regulations, yet the regulatory agency's objectives are the same as yours: they want to protect the public from dangerous drugs and worthless devices; you want to detect and avoid problems before your products go to market and shield your firm from bottom-line destroying lawsuits.

Time spent battling with a regulatory agency is time wasted, as is time spent on an inferior product. One fundamental rule should dominate your thinking: the quicker I bring a lawsuit-free product to market, the more money my firm makes.

Moral. It is better to anticipate and plan for regulatory agency objections than to try to circumvent them.

KISS

Keep it simple when you design your study, when you submit your protocol to the regulatory agency for approval, and when you prepare your reports.

Resist all temptation to use your study as a platform for experimentation, to compare alternate methods of measurement or alternate surgical procedures, unless such methods or procedures are the primary focus of and motivation for your study.

Simple, straightforward designs simplify training, encourage uniformity of execution, and are more likely to be adhered to in a uniform manner by your investigators.

Keep your intervention simple—a pill a day is preferable to three or four if you want to ensure patient compliance.

I hold little hope for a recently launched trial of a smoking cure which requires each subject to receive a preliminary three-day course of injections, take pills twice a day for 90 days, and attend weekly counseling sessions.

The same stress on simplicity should dominate your thinking when you set up time lines for the study. Follow-up examinations need to be scheduled on a sufficiently regular basis that you can forestall dropouts and noncompliance, but not so frequently that study subjects (on whose shoulders the success of your study depends) will be annoyed.

Keeping your protocol simple will make it easier for the regulatory agency to grasp and provide fewer opportunities (handles) for objection.

The preceding rules go out the window when it is the regulatory agency that insists on the complications (e.g., see Freedman et al. 1995). Ditto, if they ask for more data or more observations than you think is necessary. You can't control *them*. But, as they say in self-help psychology manuals, you can control yourself.

Keep the forms simple—resist the temptation to ask redundant questions.

Store and retrieve character data in character form. "0" and "1" may have offered space-saving advantages in the 1960s when a minicomputer provided only four thousand characters (kilobytes) of memory and tape was used for mass storage, but today when even a personal computer's memory measured in millions of characters and compact disks hold billions of bits of information, you might just as well store "Y" or "yes" for yes as convert Y to a 1. (Or was a "2" a "yes?")

Need to store the procedure's name "atherectomy?" You could code atherectomy as a 411 or 502, but why not just as "ather?" As with all things simple, eliminating codes eliminates time-wasting coding and decoding as well as a major source of error. Plus, most regulatory agencies require the data in "decoded" format.

Keep the reports simple. If you, your employees, and your spouse don't understand them, regulatory agency reviewers won't either.

PLUG THE HOLES AS THEY ARISE

Computer-assisted data entry offers a tremendous opportunity for early detection of deleterious trends resulting from discrepancies in trial design or investigator inaction. It does, that is, *if* we pay attention to and act on the information we receive.

One of the earliest steps on the U.S. path to putting a man on the moon involved putting a Rhesus monkey into orbit. The monkey was watched and probed intently and studied from every angle. Evenings and weekends, in the absence of human observers, a dozen or more recorders kept track of the monkey's temperature, blood pressure, and other vital readings. Only no one looked at the output. The monkey was in orbit before the earth-bound observers noticed he was running a fever. Yes, the monkey died.

A typical set of clinical trials today costs what that wasted spaceshot did in the 1950s. We needn't make the same mistakes that were made back then. Find the discrepancies and take advantage of the immediate availability of information that computer-assisted data entry provides to plug the holes as they arise.

PAY FOR RESULTS, NOT INTENTIONS

The most expensive single item in any study today is the physician's fee. Well, perhaps hospital charges can be appreciable as well. But you don't pay the hospital until after the surgery is completed and the patient discharged. Similarly do not pay the physician (or specialty laboratories) until all the completed forms are in hand. (See sidebar, Chapter 14.)

PLAN, DO, THEN CHECK

Even if you follow every step of the prescription outlined in the next chapter, something is bound to go wrong, or at least, to turn out differently from the way you anticipated. A study is never completed until you have reviewed the outcome, noted your errors, and, without assigning blame, prepared for the future.

Track your costs. From the very first day, it's essential you maintain a ledger of human-hours and supplies devoted to the project.

You'll find more on these topics in the remaining chapters of this text.

Chapter 3

Prescription for Success

THE PURPOSE OF THIS CHAPTER is to provide you with an outline of our prescription for success in the design and conduct of clinical trials.

PLAN

A. Include a Pre-design Phase. Form your design team (see Chapter 4). Your team's first step should be to decide whether the study is actually worth performing and whether you are ready to go forward.

Do you have the information you need on dosage, toxicity, and cross-reactions with other commonly administered drugs? Are the details of any necessary surgical procedure(s) standardized and commonly agreed on?

Do you know which if any categories of patients should be excluded from the trials? Will the market for the drug/device/bio-engineered formulation once these patients are excluded still justify performing the trials?

Begin to track your costs. Maintain a ledger of the human-hours and supplies expended on each phase of the project.

B. Design the Trials. Start with your reports and package insert. Let them determine the data you'll need.

Specify primary measures of efficacy. Decide what end points will be used to measure them. See Chapter 5.

Specify all measures of safety and any secondary measures of efficacy. Will you use checklists of adverse events at follow-up? Ask patients to volunteer concerns? Or do both? How long will the follow-up period be? See Chapter 5.

Specify eligibility requirements. Too narrow a patent will force you to repeat the trials later and may make it difficult to recruit the necessary number of subjects. Too broad a patent may doom the success of the trials by including those unlikely to benefit from the intervention.

Specify baseline measures. Include all variables that might impact treatment outcome. See Chapter 6.

Specify design parameters as defined in Chapter 6 including all of the following:

- **Treatment you will use with the control group.**
- **Extent to which investigators will be permitted knowledge of the specific treatment each patient receives.**
- **Whether you will utilize an intent-to-treat protocol.**
- **Degree of confidence you wish to have in the final results.**
- **Sample size required.**

Be aware of regulatory requirements. Guidelines for various countries can be accessed at *http://controlled-trials.com/links/guidelines.asp*.

Put your major effort into preparing for the trials, not in repairing them. Prepare for exceptions. See Chapter 7.

Don't collect data you don't need

Store and analyze the data you do collect

DO

Steps C—F can be executed in parallel.

C. Obtain Regulatory Agency Approval for the Trials. Obtaining regulatory agency approval can be as simple as submitting a written copy of the protocol you already developed. (Government agencies being what they are, you may need to reformat the document to fit their requirements.) KISS is the operating phrase. Hopefully, simplicity was exercised in the design, along with clarity in writing the proposal. See Chapter 8.

D. Form the Implementation Team. Include a pharmacologist or manufacturing specialist who will be responsible for providing the necessary supplies. Allocate resources. Have your attorney review physician contracts. Hire documenters, lead programmer, and data manager.

E. Line up Your Panel of Physicians. See Chapter 9.

F. Develop the Data Entry and Data Management Software.

- **Decide how you will collect and store the data.**

- **Decide what development and data management software you will use.**
- **Prepare a timeline for development and construction of the database and hire the necessary programming staff.**
- **Finalize the data to be collected. Determine the range of acceptable values for each individual data item. Establish the data entry formats.**
- **Develop data entry screens in sets corresponding to the individuals who will complete them.**

See Chapters 10 and 11.

G. Test the Software. Conduct both automated and ad hoc tests, the latter employing individuals who will actually use the software. See Chapter 10.

Steps H–J can be done together.

H. Train. Three topics should be covered in a training program for the investigators and their staffs. See Chapter 10.

1. **Details of the intervention. The procedures manual developed in Chapter 8 will serve as text.**
2. **Data entry**
3. **Ensuring patient compliance**

I. Recruit Patients. Recruit and enroll patients and put in place measures to monitor and ensure patient compliance. See Chapter 9.

J. Set up External Review Committees. Their composition is considered in Chapter 4 and their functions in Chapters 4 and 14.

Steps K and L can be done together.

K. Conduct the Trials.

- **Review checklist. See Chapter 12.**
- **Maintain a database and provide for its security. See Chapter 11.**
- **Maintain a schedule of regular visits to the investigators (in parallel with L). See Chapter 13.**
- **Collate data (in parallel with L). See Chapter 14.**
- **Prepare and review interim reports. Follow up on discrepancies and missing values immediately. See Chapter 14.**
- **Call meetings of the safety committee if necessitated by adverse event reports.**
- **Pay physicians and testing laboratories as completed reports are received.**

L. Develop a Suite of Programs for Use in Data Analysis. See Chapter 15.

M. Analyze and Interpret the Data. See Chapter 15.

CHECK

N. Complete the Submission.

Prepare final report to regulatory agency. See Chapter 8.

Review study with regard both to study weaknesses and to findings that may serve as the basis for future studies.

Prepare an After-Action Review. See Chapter 16.

Review the cost ledger with a view toward preparing budgets for future studies.

Check with marketing regarding preparation of journal articles, physician guides, and the like.

Begin long-term follow-up and collection of postmarketing adverse event data.

Chapter 4

Staffing for Success

YOUR FIRST STEP IN EMBARKING on a new clinical study is to staff up to meet your needs. While the natural temptation is to use those who assisted you in the past, a new approach may require new personnel with a different set of skills. The purpose of the present chapter is to list the personnel and associated skill sets you'll need to fulfill each step of the prescription outlined in the preceding chapter.

DESIGN TEAM

Given our emphasis on objectives, it should come as no surprise that the people you'll need most at the start of a project are those who will be present at the end to analyze and interpret the results.

I don't recommend the hiring of "design" experts unless they are experts at facilitating group discussions. Those who will reap are those who must sow. Nor do I advise your using someone just because they are "available." To be effective, the members of the design team must be matched to the required skill sets that we cover at length below.

These individuals include the following:

The **project manager** whose chief skill is that of a facilitator, possessing the ability to draw out and motivate others, encourage differing points of view yet obtain consensus, assign and organize tasks, and make, not defer decisions. Procrastinators need not apply.

Two physicians, one to concentrate on measures of efficacy, the other on safety. Both should be specialists in the area under investigation. As the two are intended to provide differing and, sometimes, conflicting points of view, they cannot be in a mentor-student or a supervisor-employee relationship. Both will be expected to interpret final results and sign off on reports to the regulatory agencies. One or both will serve as **medical monitor**s during the course of the trials.[4]

As discussed in the next chapter, they will be expected to provide assistance in determining what information is to be collected and how measurements are to be made and interpreted. They help in developing procedure manuals. They also will be expected to provide assistance and perhaps some direction in recruiting investigators for the study.

The medical monitors will answer all questions from investigators as to the procedures to be followed and will investigate possible protocol violations.

A **statistician**—preferably at the Ph.D. level—will participate in the development of interim reports (see Chapters 8 and 14) and will supervise the final analysis.[5] In the design phase he or she will be responsible for restatement of the design requirements in a form that lends itself to computerized analysis.[6]

One or more **clinical research monitors (CRMs)** will serve, along with the Medical Monitor, as the principal points of contact with study investigators and their staff. They will participate in literally all phases of the study. Monitors must like to travel and be able to remain away from home for extended periods (they will have to remain in the field for training and perhaps to see the first several patients through the trial process at each site). They must have excellent communication skills and be able to maintain emotional as well as intellectual empathy with physicians and their assistants. The responsibility of maintaining morale over the extent of a lengthy trial process (see Chapter 13) often falls on their shoulders.

Monitors must also have an attention to detail. They need to have good speaking voices as they will be responsible for the training in data entry of the study physicians and their staff. During the design phase they will be expected to acquire a knowledge of the clinical trial literature for the specialty under investigation.[7] Obviously their familiarity with past trials in the same area is a definite plus.

A **regulatory liaison** could be one of the above. The regulatory liaison's formal "role" is to interact with the regulatory agency, assuming (or, more accurately, sharing) the responsibility of interpreting the applicable regulations and ensuring the trials remain in compliance.

---

[4] If not, a third physician, preferably one employed by your company, will need to be appointed as medical monitor.

[5] See Chapter 15 for a comprehensive description of these duties.

[6] See, for example, the section on determining sample size in Chapter 6.

[7] See, for example, the bibliography at the end of Chapter 5.

A **marketing representative** can provide valuable input on desirable end points (you can't claim what you haven't established) and can aid in making the initial decision as to whether the trials are justified.

REGULATORY APPROVAL

I highly recommend that a single individual make all primary contacts with the regulatory authority. At some point a team physician may need to make contact with a physician employed by the regulatory agency or the team statistician with the regulatory statistician, but all such traffic should be arranged and directed by the regulatory liaison.

For preparing and reviewing submissions, the regulatory liaison should avail himself of the services of one or both of the physicians, the statistician, a medical writer, and the clinical monitors. He or she needn't be a gifted writer but should be able to direct the efforts of those who are. And, the regulatory liaison needs to be a careful reader.

The liaison should have the salesperson's gift to "mirror" those with whom he or she is interacting. (Balance is essential and a hard sell definitely not advisable.)

Finally, the liaison needs to have a positive attitude toward the regulatory agency. A need to outwit, circumvent, or simply oppose is a guaranteed recipe for disaster.

IMPLEMENTATION TEAM

Your implementation team will consist of a **pharmacologist and/or manufacturing specialist** who will be responsible for providing the drugs and/or devices needed for the trials, **clinical research monitors** who will train, deliver, and monitor the ongoing process, **technical writers** to prepare the detailed procedures manuals for use by the investigators, the **lead software developer** who will be responsible for developing the data entry screens as described in Chapter 10, and the **database manager** who will be responsible for maintaining the integrity of the collected data as described in Chapter 11. The qualifications for the latter two individuals are outlined in the next section.

You may also wish to add members whose primary concern is patient recruitment and retention.

DATA ENTRY SOFTWARE

Responsibility for choosing the appropriate software for data entry, data management, and statistical analysis is normally divided among the lead software engineer, the data manager, and the statistician. (Subject, of course, to corporate approval, a topic on which we wax apoplectic in Chapter 10.) The project leader may need to step in to resolve conflicts.

The **lead software developer** need not be a member of the programming team, but she must possess a general knowledge of both data entry and data management software and be able to prepare and maintain a flow or Gant chart for the development process. She bears overall responsibility for assembling the field specifications in collaboration with the clinical research monitor, and for approving the final screen designs. Ideally, she will also possess a knowledge of the statistical analysis software that will be employed later on.

A team of programmers will be needed to develop the data entry screens. They will not be working alone but in partnership with the clinical research monitors who bear the responsibility for the sequencing of questions and specifying the range of permissible answers. Programming sophistication is not as important as good interpersonal skills (particularly today when the software does so much of the detail work). As illustrated in Chapter 10, a knowledge of ergonomics is essential.

The size of the team will depend on the time lines that have been established. At least one member of the software design team should be from the testing group to ensure that quality is built in from the start.

Test the Software

Those who develop the software should be no more permitted to organize the final testing stages than a starting quarterback would be permitted to call defensive signals. For just as American football today has one team for offense and a second team for defense, so too should you have one team for development and one for testing: the two tasks require quite different mind-sets.

The testing team consists of one or more testing leads, the "formal" testing staff, and some "informal" testers. The **testing leads** are responsible for developing automated testing routines using such screen-capture utilities as WinRunner. (See the Appendix.) While the leads need to have a thorough understanding of the critical distinctions among unit, integration, and stress testing, the balance of the **formal testing team** can and ought to be relatively unskilled in computer use. Their task is to simulate the sort of errors that similarly unskilled personnel can make when the software is actually in use in the field. (Never mind that such "unskilled" personnel, physicians, nurses, and laboratory workers may have solid credentials in noncomputer areas.)

I have found it useful to use one of the brighter new additions to the staff to serve as devil's advocate from the very beginning of the process. In the final stages of testing, the clinical research monitors, project leader, staff physicians, and other members of the design team should be invited to participate.

And don't forget the hardware. The testing team will need computers over and above the ones you already have. Those slated to go into the trial physicians' offices would be ideal. (Or, if these are sacrosanct, then additional equipment should be rented for the duration.) You may also require additional support personnel to ensure that the testing team's computers will be up and running at all times.

DON'T FORGET THE HARDWARE

I once worked as a consultant to the team responsible for testing Xerox's ill-fated Globalview operating system. My job was to figure out why the group was falling behind. They'd doubled the number of testers, yet there was no corresponding increase in productivity. The not particularly complicated explanation was that the 12 testing personnel had only been assigned 3 computers, at least one of which was always unavailable while one or the other of the developers tried to figure out what had gone wrong.

I'm sure Xerox's middle management already knew this; they just wanted the explanation to come from an outsider. Moral: Don't just hire people; create a working environment with all the software, hardware, and other tools these people will need.

INVESTIGATOR PANELS

Putting together a panel of physicians and specialist laboratories is a nontrivial task to which (along with patient recruitment) we devote Chapter 9. Primary responsibility normally falls to one or the other of your lead physician investigators if he or she is an employee. Otherwise, recruitment becomes the project leader's responsibility. In either case, you may expect to require substantial assistance from the clinical monitors who will need to inspect each site before approval is given.

An **attorney** is needed to draft contracts with the physicians and laboratories you've recruited. (See Chapter 7.)

SITE COORDINATORS

Regardless of their prior experience with clinical trials, physicians invariably *underestimate* the amount of effort data collection will entail. Left to their own devices, physicians can and will assign supervision to overworked residents and nurses. The result, as one might expect, is both increased turnover of personnel and a degradation in the quality of the data that is collected. The smart drug or device company will pay all or most of the salary of a *coordinator* at each site, thus ensuring both quality and continuity.

Moreover the medical records of most physician offices today are paper-oriented. A trained coordinator onsite is essential to facilitate the conversion to direct data entry, to ensure that computers are placed where they lend themselves to immediate use.

The coordinator, usually a nurse, is responsible for seeing that data are entered in a timely fashion and for ensuring prompt transmission of the data to the trial's sponsor. She will unpack the drugs and devices and verify that they are as requested. Often she will take on the responsibility of administering informed consent to the prospective patients. If not, she becomes responsible for seeing that informed consent is administered. If ambiguities arise or if problems occur in any aspect of the study (including the software and the hardware), she is responsible for notifying the clinical research monitor.[8]

EXTERNAL REVIEW COMMITTEES

You will need to establish several external committees whose members are independent of both your investigators and your staff to review trial findings. Each external committee will serve one of three main functions:

1. **To review measures (e.g., X rays, EEGs, ECGs, and angiograms) whose interpretation can be subjective.**
2. **To determine whether adverse events should be deemed intervention related.**
3. **To determine whether to continue or modify the trials based on interim findings.**

---

[8] The presence of a site coordinator does not relieve the CRM of the necessity of making on-site inspections. See Chapter 14.

The committees appointed to review and interpret angiograms, X rays, EEGs, and ECGs. should consist of experts in the specific diagnostic area. The assignment-of-cause committee would consist of specialists in the disease process. The trial modification committee should include a statistician as well as physicians. All the committees should be able to call on additional experts—bioengineers, epidemiologists, geneticists, or pharmacologists—whenever they feel such services are warranted.

The primary duty of committee members is to ensure the safety of the participants in the trials. Their secondary responsibility is to ensure the integrity of the trials: the investigators, the regulatory agency, and the sponsor will rely on their advice. It is essential that membership be composed of individuals who are already recognized experts in their fields, and that these individuals lack any other direct connection with the sponsor of the trials.

A member of the design team should serve as a permanent liaison with each committee.

Recruit and Enroll Patients

Without outside assistance, the typical panel physician may recruit as few as one-fifth of the patients he or she promised to deliver originally. Some may recruit none, taking their setup money and contributing nothing more, or recruit only one patient so that there is no offsetting control. We've outlined a number of techniques for increasing recruitment in Chapter 9. An experienced CRM could be placed in charge of the overall recruiting effort, aided by sponsor-paid site coordinators. Or you may find it more expedient to rely on the service of a professional recruiter. In any event, I would not recommend you wait "to see how the numbers turn out."

Conduct the Trials

Principal responsibility for the actual conduct of the trials rests with the project leader, database manager, and clinical research monitors. The latter's task is greatly simplified if you make use of paid study coordinators at each treatment site. The statistician may be needed to assist in the preparation of interim reports. Indeed, depending on the nature and duration of the study, virtually all members of the design and implementation teams may be called upon. In Chapter 14 we discuss the need for assembling **external review panels** and their recommended composition.

The project leader is responsible for authorizing payment to study physicians and other contract resource personnel (pharmacologists, radiologists, testing laboratories) as each individual milestone is completed. Approval is generally pro forma once the clinical research monitors report completion.

The **database manager** has the continuing responsibility of seeing that the data are stored correctly, that data integrity are maintained and that data are readily retrievable. His or her knowledge must extend beyond an understanding of the data management software to security (maintaining on-site and off-site backups) and quality control.

PROGRAMS FOR DATA ANALYSIS

Development of programs for data analysis should be started on or before the actual beginning of the trials. One or more **statistical programmers** will work under the direction of the statistician. The ideal statistical programmer will also be a member of the team that develops the data entry screens. (Candidly, programmers who possess the dual set of skills are in extremely short supply.)

YOU DON'T NEED TO DO IT YOURSELF BUT YOU PROBABLY SHOULD

"The editors [of 13 leading medical journals] will criticize pharmaceuticals companies for their use of private, nonacademic research groups—called "contract research organizations' (CRO's) instead of scientists connected to universities and hospitals.

"CRO's fail to provide sufficient oversight of clinical trials."
Financial Times, **10 September 2001**

You could use your full-time employees to conduct the clinical trials, or you could hire, on a temporary basis, a few or all of the people you need. You could have a staffing firm supply the programmers you need, and place an advertisement for a consulting statistician. You could even hire a contract research organization (CRO) that will design and conduct part or all of the trials for you.

If you're a struggling one-product start-up with just enough working capital, the lease versus purchase option can look awfully attractive. Even well-established firms occasionally farm out their trials when they view their expanded needs as temporary or when a hiring freeze would make the trials impossible to conduct otherwise.

But a large firm doesn't just say to a CRO, "here's some money, go do the trials, get back to me when you have the results." They establish milestones similar to those you encounter chapter by chapter in this text. And an experienced full-time employee rides herd over the CRO's efforts.

Whether you decide to hire full-time employees, lease contract employees, or farm out the study to a contract research organization, you continue to bear the responsibility for the successful conduct and administration of the trials. Read on.

Analyze and interpret the data. Though the analysis is primarily the statistician's responsibility, he will need to work through the clinical research monitors and with the database managers to resolve any remaining issues of interpretation and data discrepancies. Clinical significance may be quite different from statistical significance (a point considered at length in Chapter 15); interpretation of trial results becomes the responsibility of the project leader drawing on the expertise of all the members of the design team.

Who Is on the Team?

	Number	*Reports to*	*Chief Roles*
Project leader (PL)	**1**		**Facilitator, manager**
Clinical research monitor (CRM)	**1–2**	**PL**	**Liaison with investigators; assist PL with project administration**
Site coordinators	**1 per site**	**CRM**	**Coordinate activities on site; determine data to be collected**
Physicians	**2**		**Monitor safety and efficacy**
Pharmacologist or manufacturing specialist	**1**		**Prepare and deliver materials**
Statistician (ST)	**1**		**Determine sample size and methods of analysis**
Regulatory liaison	**1**		**Liaison with regulatory agencies**
Marketing liaison	**1**		**Needed during design/ analysis**
Attorney	**1**		**Draft investigator contracts**
Technical writers	**1+**		**Prepare investigator manuals**
Lead programmer (LP)	**1**		**Develop data entry screens**
Programmers	**2+**	**LP**	**Develop data entry screens and programs for data extraction**
Database manager	**1**	**PL**	**Maintain integrity and security of data**
Programmer, statistical	**1**	**ST**	**Develop programs for data extraction, data analysis, and reports.**

To say nothing of investigators, investigational laboratories, safety and efficacy review committees, and patients.

THE PEOPLE YOU DON'T NEED

Too often downsizing is accomplished on an across-the-board percentage basis. Each manager is given a quota, say 8%, to eliminate, and then left to his or her own devices. A no-brainer that hardly justifies the high incentives paid corporate cost cutters. True downsizing means reorganization and the redrawing of job descriptions to ensure effective performance.[9] In other words, someone who is not part of the design team and who will not be able to assist at other stages in the conduct of the study isn't needed any more. Put them to work elsewhere in the company, retrain them, or let them go.

More than one manager has told me, "If I do as you suggest and my head count drops below a certain number, then they're liable to let *me* go." If you're that manager's manager or, more aptly, the executive who dreamed up this absurd head-count policy, do your company and your own stock options a favor: resign.

---

[9] Don't attribute this quote to me, Professor Deming and the pioneer industrial engineers of the 1920s said it long before.

Chapter 5

Design Decisions

From the outset of the study, we are confronted with the need to make a large number of decisions, including, not least, "should the study be performed?" A clinical trial necessitates a large financial investment. Once we launch the trials, we can plan on tying up both our investment and the work product of several dozen individuals for at least the next two to six years. Planning pays.

Seven major design decisions that must and should be made before the trials begin are covered in the present chapter:

1. **Should the study be performed?**
2. **What are the study's objectives?**
3. **What are the primary and secondary response variables?**
4. **How will the quality of the information be assured?**
5. **What types of subjects will be included in the study?**
6. **What is the time line of the study?**
7. **How will the study be terminated?**

Five somewhat more technical design decisions are covered in the chapter following:

1. **What experimental design will be utilized?**
2. **What baseline measurements will be made on each patient?**
3. **Will it be a single-blind or a double-blind study?**
4. **What sample size is necessary to detect the effect?**
5. **How many examination sites will we need?**

PRE-DESIGN CHECKLIST

Before you can begin full-scale clinical trials, you need to establish:

- Mutagenicity, carcinogenicity, and toxicity in animals
- Mechanism of action in humans
- Maximum tolerated dose
- Minimum effective dose

We deal in Chapter 7 with the large number of minor details that must be thought through before we can conclude our preparations.

SHOULD THE STUDY BE PERFORMED?

We should always hesitate to undertake extensive trials when a surgical procedure is still in the experimental stages, or when the cross-effects with other commonly used drugs are not well understood. A cholesterol-lowering agent might well interfere with a beta blocker, for example.

If your study team is still uncertain about the intervention's mode of action, it may be advisable to defer full-scale trials till a year or so in the future and perform instead a trial of more limited scope with a smaller, more narrowly defined study population. For example, you might limit your trial to male nonsmokers between 20 and 40 who are not responding to current medications.

No full-scale long-term clinical trials of a drug should be attempted until you have first established both the maximum tolerable dose and the anticipated minimum effective dose. (In the United States, these are referred to as Phase I and Phase II clinical trials, respectively.) You should also have some ideas concerning the potential side effects.[10]

ONE TRIAL? OR MANY?

A single large-scale trial might appear more cost effective in the short term, says Michael Chernick of Novo Nordisk, but multiple tightly focused clinical trials generally are cleaner and faster. Multiple trials might be preferable in the following circumstances:

- Testing for different disease conditions
- Testing in different subpopulations
- Testing for different effects
- Monotherapy in one trial, combination therapy in another
- Different control groups for one-on-one comparisons for different benefits

The trials need not be concurrent and can often benefit from the results of other trials in their final design.

STUDY OBJECTIVES

I'm constantly amazed by the number of studies that proceed well into the clinical phase without any clear-cut statement of objectives. The executive committee has decreed "the intervention be taken to market" and this decree is passed down the chain of command without a single middle manager bothering or daring to give the decree a precise written form.

---

[10] See Fazzari, Heller, and Scher (2000).

Begin by stating your principal hypothesis such as:

- **An increase in efficacy with no increase in side effects**
- **A decrease in side effects with no decline in efficacy**
- **No worse than but less costly and/or less invasive**

For Motrin™, for example, the principal hypothesis was that Motrin would provide the same anti-inflammatory effects as aspirin without the intestinal bleeding that so often accompanies continued aspirin use.

Keep the package insert in mind. For naproxen, another anti-inflammatory, the package insert reads: "In patients with osteoarthritus, the therapeutic action of naproxin has been shown by a reduction in join pain or tenderness, an increase in range of motion in knee joints, increased mobility as demonstrated by a reduction in walking time, and improvement in capacity to perform activities of daily living impaired by the disease.

"In clinical studies . . . naproxin has been shown to be comparable to aspirin and indomethacin . . . but the frequency and severity of the milder gastrointestinal adverse effects . . . and nervous system adverse effects were less in naproxin treated patients than in those treated with aspirin and indomethacin."

SET UP A DEFENSIVE TEAM

From the very start of the project, you need to establish a group whose primary purpose is to find the holes left in your design. I suggest a group rather than an individual because in today's corporate environment we all want to be thought of as team players. Moreover not everyone makes an effective critic. If you are managing several projects simultaneously, then the members of one study group may be called on to criticize the efforts of the other. Otherwise, and in particular if your firm is a small one, it may be best to call on external consultants. Of course, your own role should be that of a facilitator rather than a proponent of any specific point of view.

The objectives of your study should be stated as precisely as possible. Consider the following: "The purpose of this trial is to demonstrate that X763 is as effective as aspirin in treating stress-induced headaches and has fewer side effects."

Not very precise, is it? Here is a somewhat more informative alternative: "The purpose of this trial is to demonstrate that in treating stress-induced headaches in adults a five-grain tablet of X763 is as effective as two five-grain tablets of aspirin and has fewer side effects." This is a marked improvement, though it is clear we still need to define what we mean by "effective."

A more general statement of objectives that may be used as template for your own studies takes the following form. "The purpose of this trial is to demonstrate that:

- **in treating conditions A, B, C**
- **with subjects having characteristics D, E, F**
- **an intervention of the form G**
- **is equivalent to/ as effective as/ as or more effective than an intervention of the form H**
- **and has fewer side effects."**

Again, we still need to define what we mean by "effective" and to list some if not all of the side effects we hope to diminish or eliminate.

PRIMARY END POINTS

Our next task is to determine the primary end points that will be used to assess efficacy. Here are a few guidelines:

- **Objective criteria are always preferable to subjective.**
- **True end points such as death or incidence of strokes should be employed rather than surrogate response variables such as tumor size or blood pressure. The latter is only appropriate (though not always avoidable) during the early stages of clinical investigation when trials are of short duration.**
- **The fewer the end points the better. A single primary end point is always to be preferred as it eliminates the possibility that different end points will point in different directions. On the other hand, as we will see in Chapter 14 on data analysis, sometimes more effective use of the data can be made using a constellation of well-defined results.**

The obvious exceptions are when (1) surrogate end points are employed and a change in a single factor would not be conclusive, (2) your marketing department hopes to make multiple claims, (3) competing products already make multiple claims.

The end point can be determined in two ways:

1. **Duration of the symptom or disease.**
2. **Severity of the symptom or disease at some fixed point after the start of treatment. This latter can be expressed either in terms of (a) a mean value or (b) the proportion of individuals in the study population whose severity lies below some predetermined fixed value.**

END POINT OR SURROGATE?

I'm taking drugs currently to control my blood pressure and to lower my cholesterol. Thus my interests will be served if my diastolic blood pressure remains below 90 and my cholesterol dips below 200. Or will they? As my passion for ice-cream reveals, I don't really care about cholesterol at all, or at least I didn't for most of my life. But I do not want to have a heart attack or a stroke and I've been told that if I keep my blood pressure down and my cholesterol levels low I may well avoid both.

It is both less time-consuming and less expensive to measure changes in surrogate variables like cholesterol and blood pressure than it is to track survival. The former can be detected in days to weeks; the latter will (hopefully in my case) take many years. But can we always be sure that the surrogate variable we measure is directly related to the end point that is our real interest?

Because very large-scale, very long-term clinical trials were conducted with government support, clinical trials employing surrogate variables such as cholesterol as end points are acceptable in some areas. But not in all. There are many documented reports of surrogate variables that have failed abysmally as predictors of sudden cardiac death (CAST, 1989), cancer survival (Fleming, 1995), or AIDS recovery (Fleming, 1995).

Any attempt to use a surrogate variable is sure to be viewed skeptically by the regulatory agency. It was not until well after the completion of LifeCore's clinical trials of its Intergel^{TM} adhesion prevention solution, that adhesion was declared to be an end point rather than a surrogate.

On a further practical level, you cannot advertise what you do not demonstrate, and a failure to use actual end points will limit your subsequent marketing claims.

For a blood-pressure lowering agent such as metoprolol, the primary end point is diastolic blood pressure. For an anti-inflammatory such as Motrin, it might be either the duration or the extent of the inflammation. For a coronary-stenosis reducing surgical procedure or device, it might be the percentage of stenosis or the percentage of the population with less than 50% stenosis (termed "binary restenosis").

An exact quantitative definition should be provided for each end point. You also will need to specify how the determination will be made and who will make it. Subjective? Objective? By the treating physician? Or by an independent testing laboratory? Is the baseline measurement to be made before or after surgery?

In a study of several devices for maintaining flow through coronary arteries, the surgeon who performed the operation made the initial determination of stenosis. But it was decided that the more accurate and “official” reading would be made from an angiogram by an independent laboratory.

How much give in dates is permitted?—patients have been known not to appear as scheduled for follow-up exams. What if a patient dies during the study or requires a further remedial operation? How is the end point of such a patient to be defined?

Don’t put these decisions off till some later date; make them now and make them in writing lest you risk not collecting the data you will ultimately need.

Secondary End Points

Secondary[11] end points are used most often to appraise the *safety* of an intervention.

For a blood-pressure lowering agent like metoprolol these might include dizziness and diarrhea. But the systolic blood pressure would also be of interest.

For an anti-inflammatory, the most important are intestinal bleeding and ulcers. How does one detect and measure intestinal bleeding? Two ways, by self-evaluation and by measuring the amount of blood in the stool. Data relating to both must be collected.

For a coronary-stenosis reducing surgical procedure or device, the primary concern is with other procedure- and condition-related adverse events including death, myocardial infarctions, and restenosis severe enough to require further operations.

To ensure that you will collect all the data you need, a careful review of past clinical and pre-clinical experience with the present and related interventions is essential. For example, suppose that extremely high doses of your new agent had resulted in the presence of abnormal blood cells in mice. While such abnormal cells may be unlikely at the therapeutic dose you are using in the trials, to be on the safe side, blood tests should be incorporated in the trial’s follow-up procedure.

Don’t Collect Data You Don’t Need

Store and Analyze the Data You Do Collect

During the trial and afterward, you will probably want to record the frequency of all adverse events, of specific adverse events, and of

---

[11] The use of the terms “primary” and “secondary” can be misleading. Quite often in long-term clinical trials we are already confident in the efficacy of a treatment but are extending the duration of the trial so that we can be equally certain of the absence of long-term negative effects.

CHECKLIST OF MEASUREMENTS

- What is the nature of your intervention?
  - How will it be administered?
  - What is its duration?
- You are planning to test for efficacy.
  - What are your primary end points?
  - When will the measurements be made?
  - How will the measurements be made?
    - Who will make them?
    - What units will be used?
  - Who will interpret the measurements?
  - What quantitative results do you expect?
- You must test for safety.
  - What short-term side effects are expected?
  - How do you plan to measure them?
  - What quantitative results do you expect?
  - How soon can you expect to observe them?
  - What long-term side effects are expected?
  - How do you plan to measure them?
  - What quantitative results do you expect per 100 patients?

those events directly related to the intervention that exceed a certain level of severity.

You should also determine *how* the adverse event data are to be collected. By use of a checklist—“Since your last appointment, did you experience fever? nausea? dizziness?” Or a volunteered response—“Have you had any problems since your last visit?” Elicited responses tend to yield a higher frequency of complaints. To be on the safe side, use both methods. Of course, hospitalizations, emergency treatment, and phoned-in complaints between visits must always be recorded.

Some secondary end points may also concern efficacy. For example, in a study of sedatives, you might be interested in how rapidly the patient obtained relief.

Tertiary End Points. Tertiary end points such as costs may or may not be essential to your study. Don’t collect data you don’t need. When in doubt, let your marketing department be your guide.

BASELINE DATA

You will need to specify what baseline data should be gathered prior to the start of intervention and how it will be gathered—by interview, questionnaire, physical examination, specialized examinations (angiograms, ultrasound, MRI), and/or laboratory tests. Baseline data will be used both to determine eligibility and, as discussed in the next chapter, to stratify the patients into more homogeneous subgroups.

Be comprehensive. Unexpected differences in outcome (or lack thereof) may be the result of differences in baseline variables. What isn't measured can't be accounted for.

WHO WILL COLLECT THE DATA?

One further step involves grouping the questions in accordance with the individual who will be entering the data, for example, demographics and risk factors by the interview nurse with review by the physician, and laboratory results by the lab itself or by the individual who receives the report. These groupings will form the basis for programming the case report forms (see Chapter 10).

Finally, I would recommend you charge specific individuals with the responsibility of addressing each of the points raised in the preceding sections. The design committee can then function as a committee should in reviewing work that has already been performed.

QUALITY CONTROL

The secret of successful clinical trials lies in maintaining the quality of the data you collect. The most frequent sources of error are the following:

- **Protocol deviations that result when the intervention is not performed/administered as specified**
- **Noncompliance of patients with the treatment regimen**
- **Improperly labeled formulations**
- **Improperly made observations**
    - **Inaccurate measuring devices**
    - **Inconsistent methods of observation, the result of**
        - **Ambiguous directions**
        - **Site-to-site variation**
        - **Time-period to time-period variation**
    - **Fraud (sometimes laziness, sometimes a misguided desire to please)**
- **Improperly entered data**
- **Improperly stored data**

Among the more obvious preventive measures are the following:

1. **Keep the intervention simple. I am currently serving as a statistician on a set of trials where, over my loudest protests, each**

patient will receive injections for three days, self-administer a drug for six months, and attend first semiweekly and then weekly counseling sessions over the same period. How likely are these patients to comply?

2. **Keep the experimental design simple (see Chapter 6).**
3. **Keep the data collected to a minimum.**
4. **Pretest all questionnaires to detect ambiguities.**
5. **Use computer-assisted data entry to catch and correct data entry errors as they are made (see Chapter 10).**
6. **Ensure the integrity and security of the stored data (see Chapter 11).**
7. **Prepare a highly detailed procedures manual for the investigators and investigational laboratories to ensure uniformity in treatment and in measurement. Provide a training program for the investigators with the same end in mind.**

   This manual should include precise written instructions for measuring each primary and secondary end point. It should also specify *how* the data are to be collected. For example, are data on current symptoms to be recorded by a member of the investigator's staff, or by the self-administering patient?
8. **Monitor the data and the data collection process. Perform frequent on-site audits. In one series of exceptionally poorly done studies Weiss et al. (2000) uncovered the following flaws:**
   - **Disparity between the reviewed records and the data presented at two international meetings**
   - **No signed informed consent**
   - **No record of approval for the investigational therapy**
   - **Control regimen not as described in the protocol**
9. **Inspect the site where the drugs or devices are packaged; specify the allowable tolerances; repackage or relabel drugs at the pharmacy so that both the patient's name and the code number appear on the label; draw random samples from the delivered formulations and have these samples tested for potency at intervals by an independent laboratory.**
10. **Write and rewrite a patient manual to be given to each patient by their physician. Encourage and pay investigators to spend quality time with each patient. Other measures for reducing dropouts and ensuring patient compliance are discussed in Chapter 9.**

STUDY POPULATION

Your next immediate question is how broad a patent to claim. That is, for what group of patients and for what disease conditions do you feel your intervention is appropriate?

Too narrow a claim may force you to undertake a set of near duplicate trials at a later date. Too broad a claim may result in withdrawal of the petition for regulatory approval simply because the treatment/device is inappropriate for one or more of the subgroups in the study (e.g., infants and pregnant women). This decision must be made at the design stage.

Be sure to have in hand a list of potential contra-indictions based on the drug's mechanism of action as well as a list of common medications with which yours might interact. For example, many lipid-lowering therapies are known to act via the liver, and individuals with active liver disease are specifically excluded from using them. Individuals using erythromycin or oral contraceptives might also have problems. If uncertain about your own procedure, check the package inserts of related therapies.

Eligibility requirements should be as loose as possible to ensure that an adequate number of individuals will be available during the proposed study period. Nonetheless, your requirements should exclude all individuals

- **Who might be harmed by the drug/device**
- **Who are not likely to comply with the protocol**
- **For whom the risks outweigh any possible benefits**

Obviously there are other protocol-specific criteria such as concurrent medication that might call for exclusion of a specific patient.

Generally, the process of establishing eligibility requirements like that of establishing the breadth of the claim is one of give and take. The emphasis of the "give" being to recruit as many patients as possible, the "take" being based on the recognition that there is little point in recruiting patients into a study who are unlikely to make a positive contribution to the end result.

As well as making recruitment difficult—in many cases a pool of 100 potential subjects may yield only 2 or 3 qualified participants—long lists of exclusions also reduce the possibility of examining treatment responses for heterogeneity, a fact that raises the issue of generalization of results (e.g., see Keith, 2001).

BEGIN WITH YOUR REPORTS

Imagine you are doing a trial of cardiac interventions. A small proportion of patients have more than one diseased vessel. Would you:

- Report the results for each vessel separately?
- Report the results on a patient-by-patient basis, choosing one vessel as representative? using the average of the results for the individual vessels?
- Restrict the study to patients with only a single diseased epicardial vessel?

In limiting your claims, be precise. For example, "exclude all those with diastolic blood pressure over

```
E-BAS-F-F------F------F-------------------F-------T
```

FIGURE 5.1 Trial Time Line Example. (E) Eligibility determination and initial baseline measurements; (B) baseline measurements; (A) assignment to treatment; (S) start of intervention; (F) follow-up exam; (T) final follow-up exam and termination of trial. Time scale in weeks.

105 mm Hg as measured on two occasions at least one week apart." A less precise statement, such as "exclude those with severe hypertension" is not adequate and would be a future source of confusion.

Though your ultimate decision must, of necessity, be somewhat arbitrary, remember that a study may always be viewed as one of a series. Though it may not be possible to reach a final conclusion (at least one acceptable to the regulatory agency) until all the data are in, there may be sufficient evidence at an earlier stage to launch a second broader set of trials before the first set has ended.

TIMING

Your next step is to prepare a time line for your trials as shown in Figure 5.1, noting the intervals between the following events:

- **Determination of eligibility**
- **Baseline measurement**
- **Treatment assignment**
- **Beginning of intervention**
- **(If applicable) Release from hospital**
- **First and subsequent follow-ups**
- **Termination.**

Baseline observations that could be used to stratify the patient population should be taken at the time of the initial eligibility exam. (See the next chapter for a more complete explanation.) The balance of the baseline measurements should be delayed until just before the beginning of intervention, lest there be a change in patients' behavior. Such changes are not uncommon, as patients, beginning to think of themselves as part of a study, tend to become more health conscious.

Follow-up examinations need to be scheduled on a sufficiently regular basis that you can forestall dropouts and noncompliance, but not so frequently that study subjects (on whose shoulders the success of your study depends) will be annoyed.

CLOSURE

You also need to decide now and document how you plan to bring closure to the trials. Will you follow each participant for a fixed period? Or will you terminate the follow-up of all participants on a single fixed date? What if midway through the trials, you realize your drug/device poses an unexpected risk to the patient? Or (hopefully) that your drug/device offers such advantages over the standard treatment that it would be unethical to continue to deny control patients the same advantages. We consider planned and unplanned closure in what follows.

Planned Closure

Enrollment can stretch out over a period of several months to several years. If each participant in a clinical trial is followed for a fixed period, the closeout phase will be a lengthy one, also. You'll run the risk that patients who are still in the study will break the treatment code. You'll be paying the fixed costs of extended monitoring even though there are fewer and fewer patients to justify the expenditure.

TABLE 5.1 Comparison of Closeout Policies

	Enrollment Phase	Closeout	Total
Fixed term	9 months	12 months	21 months
Fixed date	9 months	12 to 21 months	21 months

WHO WILL DO THE MONITORING?

Monitoring for quality control purposes will be performed by a member of your staff, as will monitoring for an unusual frequency of adverse events. But at certain intermediate points in the study, you may wish to crack the treatment code to see if the study is progressing as you hoped. Cracking the code may also be mandated if there have been an unusual number of adverse events. If a member of your staff is to crack the code, she should be isolated from the investigators in order not to influence them with the findings. The CRM should not be permitted to crack the code for this very reason.

One possibility is to have an independent panel make the initial and only review of the decoded data while the trials are in progress. Greenberg Report (1988) and Fleming and DeMets (1993) have offered strong arguments for this approach, while Harrington et al. (1994) have provided equally strong arguments against.

Our own view is that a member of your staff should perform the initial monitoring but that modification or termination of the trials should not take place until an independent panel has reviewed the findings. (Panel members would include experts in the field of investigation and a statistician.)

And you'll still be obligated to track down each patient once all the data are in and analyzed in order for their physicians to give them a final briefing.

By having all trials terminate on a fixed date, you eliminate these disadvantages while gaining additional, if limited, information on long-term effects. The fixed date method is to be preferred in cases when the study requires a large number of treatment sites.

Unplanned Closure

A major advantage of computer-assisted direct data entry is that it facilitates monitoring the results to obtain early indications of the success or failure of the drug or device that is under test. (See Chapter 14.) Tumors regress, Alzheimer's patients become and stay coherent, and six recipients of your new analgesic get severe stomach cramps. You crack the treatment code and determine that the results favor one treatment over the other. Or, perhaps, that there is so little

BEWARE OF HOLES IN THE INSTRUCTIONS

The instructions for Bumbling Pharmaceutical's latest set of trials seemed almost letter perfect. At least they were lengthy and complicated enough that they intimidated anyone who took the time to read them. Consider the following, for example:

"All patients will have follow-up angiography at eight ±0.5 months after their index procedure. Any symptomatic patient will have follow-up angiograms any time it is clinically indicated. In the event that repeat angiography demonstrates restenosis in association with objective evidence of recurrent ischemia between zero and six months, that angiogram will be analyzed as the follow-up angiogram. An angiogram performed for any reason that doesn't show restenosis will qualify as a follow-up angiogram only if it is performed at least four months after the index intervention.

"In some cases, recurrent ischemia may develop within 14 days after the procedure. If angiography demonstrates a significant residual stenosis (>50%) and if further intervention is performed, the patient will still be included in the follow-up analyses that measure restenosis."

Now, that's comprehensive. Isn't it? Just a couple of questions: If a patient doesn't show up for their eight-month follow-up exam, but does appear at six months and one-year, which angiogram should be used for the official reading? If a patient develops recurrent ischemia 14 days after the procedure and a further intervention is performed, do we reset the clock to zero days?

Alas, these holes in the protocol were discovered by Bumbling's staff only *after* the data were in hand and they were midway through the final statistical analysis. Have someone who thinks like a programmer (or, better still, have a computer) review the protocol before it is finalized.

difference between treatments as to fail to justify continuing the trials.[12] You have the findings confirmed by your external review panel (see sidebar). Do you and should you discontinue the trials?

One school of thought favors that you continue the trials but modify your method of allocation to treatment. If the early results suggest your treatment is far the superior, then two-thirds or even three-fourths of the patients admitted subsequently would receive your treatment, with a reduced number continuing to serve as controls (e.g., see Wei et al., 1990). Others would argue that continuing to deny the most effective treatment to *any* patient is unethical. The important thing is that you decide in advance of the trials the procedures you will follow should a situation like this arise.

If you find it is your product that appears to be causing the stomach cramps, you'll want a thorough workup on each of the complaining patients. It might be the cramps are the result of a concurrent medication; clearly, modifications to the protocol are in order. You would discontinue giving the trial medication to patients taking the concurrent medication but continue giving it to all others. You'd make the same sort of modification if you found that the negative results occurred only in women or in those living at high altitudes.

My advice. Set up an external review panel that can provide unbiased judgments.

BE DEFENSIVE. REVIEW. REWRITE. REVIEW AGAIN

The final step in the design process is to review your proposal with a critical eye. The object is to anticipate and, if possible, ward off external criticism. Members of your committee, worn out by the series of lengthy planning meetings, are usually all too willing to agree. It may be best to employ one or more reviewers who are not part of the study team. (See Chapter 8.)

Begin by reducing the protocol to written form so that gaps and errors may be readily identified. You'll need a written proposal to submit to the regulatory agency. And, as personnel come and go throughout the lengthy trial process, your written proposal may prove the sole uniting factor.

Lack of clarity in the protocol is one of the most frequent objections raised by review committees. Favalli et al. (2000) reviewed several dozen protocols looking for sources of inaccuracy. Problems in data management and a lack of clarity of the protocol and/or case report forms were the primary offenders. They pointed out that training and supervision of data managers, precision in writing protocols, standardization of the data entry process, and the use of a checklist for therapy data and treatment toxicities would have avoided many of these errors.

---

[12] See Greene et al. (1992) for other possible decisions.

Reviewing the university group diabetes program study, Feinstein (1971) found at least five significant limitations:

1. **Failure to define critical terms, such as "congestive heart failure." Are all the critical terms in your protocol defined? Or is there merely a mutual unvoiced and readily forgotten agreement as to their meaning? Leaving ambiguities to be resolved later runs the risk that you will choose to resolve the ambiguity one way and the regulatory agency another.**
2. **Vague selection criteria. Again, vagueness and ambiguity only create a basis for future disputes.**
3. **Failure to obtain important baseline data. You and your staff probably have exhausted your own resources in developing the initial list so that further brainstorming is unlikely to be productive. A search of the clinical literature is highly recommended and should be completed before you hire an additional consultant to review your proposal.**
4. **Failure to obtain quality-of-life data during trial. Your marketing department might have practical suggestions.**
5. **Failed to standardize the protocol among sites. Here is another reason for developing a detailed procedures manual. Begin now by documenting the efforts you will make through training and monitoring to ensure protocol adherence at each site.**

Other frequently observed blunders include absence of concealment of allocation in so-called blind trials, lack of justification for nonblind trials, not using a treatment for the patients in the control group, inadequate information on statistical methods, not including sample size estimation, not establishing the rules to stop the trial beforehand, and omitting the presentation of a baseline comparison of groups. These topics are covered in the next chapter.

CHECKLIST FOR DESIGN

Stage I of the design phase is completed when you've established the following:

- **Objectives of the study**
- **Scope of the study**
- **Eligibility criteria**
- **Primary and secondary endpoints**

- **Baseline data to be collected from each patient**
- **Follow-up data to be collected from each patient**
- **Who will collect each data item**
- **Time line for the trials**

Stage II of the design phase is completed when you've done the following:

- **Determined how each datum is to be measured**
- **Determined how each datum is to be recorded**
- **Grouped the data items that are to be collected by the same individual at the same time (see Chapter 10)**
- **Developed procedures for monitoring and maintaining the quality of the data**
- **Determined the necessary sample size and other aspects of the experimental design (see the next chapter)**
- **Specified how exceptions to the protocol will be handled (see Chapter 7)**

BUDGETS AND EXPENDITURES

> **"Those who will not learn from the lessons of history will be forced to repeat them."**

Begin now to track your expenditures. Assign a number to the project and have each individual who contributes to the design phase record the number of hours spent on it. See Chapter 15.

FOR FURTHER INFORMATION

A great many texts and journal articles offer advice on the design and analysis of clinical trials. We group them here into three categories:

1. **General-purpose texts**
2. **Texts that focus on the conduct of trials in specific medical areas**
3. **Journal articles**

General-Purpose Texts

Chow S-C; Liu J-P. (1998). *Design and Analysis of Clinical Trials: Concept and Methodologies*. New York: Wiley.

Cocchetto DM; Nardi RV. (1992). *Managing the Clinical Drug Development Process*. New York: Dekker.

Friedman LM; Furberg CD; DeMets DL. (1996). *Fundamentals of Clinical Trials*, 3rd ed. St. Louis: Mosby.

Iber FL; Riley WA; Murray PJ. (1987). *Conducting Clinical Trials*. New York: Plenum Medical.

Mulay M. (2001). *A Step-by-Step Guide to Clinical Trials.* Sudbury, MA: Jones and Bartlett.

Spilker B. (1991). *Guide to Clinical Trials*. New York: Raven Press.

Texts Focusing on Specific Clinical Areas

Goldman DP, et al. (2000). *The Cost of Cancer Treatment Study's Design and Methods.* Santa Monica, CA: Rand.

Kertes PJ; Conway MD, eds. (1998). *Clinical Trials in Ophthalmology: A Summary and Practice Guide.* Baltimore: Williams and Wilkins.

Kloner RA; Birnbaum Y; eds. (1996). *Cardiovascular Trials Review.* Greenwich, CT: Le Jacq Communications.

Max MB; Portenoy RK; Laska EM. (1991). *The Design of Analgesic Clinical Trials*. New York: Raven Press.

National Cancer Institute. (1999). *Clinical Trials: A Blueprint for the Future*. Bethesda, MD: National Institutes of Health.

Paoletti LC; McInnes PM, eds. (1999). *Vaccines, From Concept to Clinic: A Guide to the Development and Clinical Testing of Vaccines for Human Use*. Boca Raton, FL: CRC Press.

Pitt B; Desmond J; Pocock S. (1997). *Clinical Trials in Cardiology*. Philadelphia: Saunders.

Prien RF; Robinson DS, eds. (1994). *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines / in association with the NIMH and the ACNP*; New York: Raven Press.

Journal Articles

The following journal articles provide more detailed analyses and background of some of the points considered in this chapter.

CAST (Cardiac Arrhythmia Suppression Trial). (1989). Investigators: Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N. Engl. J. Med.* 321:406–412.

Ebi O. (1997). Implementation of new Japanese GCP and the quality of clinical trials—From the standpoint of the pharmaceutical industry. *Gan To Kagaku Ryoho* 24:1883–1891.

Favalli G; Vermorken JB; Vantongelen K; Renard J; Van Oosterom AT; Pecorelli S. (2000). Quality control in multicentric clinical trials: An experience of the EORTC Gynecological Cancer Cooperative Group. *Eur. J. Cancer* 36:1125–1133.

Fazzari M; Heller G; Scher HI. (2000). The phase II/III transition: Toward the proof of efficacy in cancer clinical trials. *Control. Clin. Trials* 21:360–368.

Fleming TR. (1995). Surrograte markers in AIDS and cancer trials. *Stat. Med.* 13:1423–1435.

Fleming TR; DeMets DL. (1993). Monitoring of clinical trials: issues and recommendations. *Control. Clin. Trials* 14:183–197.

Greene HL; Roden DM; Katz RJ, et al. (1992). The cardiac arrhythmia suppression trial: First CAST . . . then CAST II. *J. Am. Coll. Cardiol.* 19:894–898.

Harrington D; Crowley J; George SL; Pajak T; Redmond C; Wieand HS. (1994). The case against independent monitoring committees. *Statist. Med.* 13:1411–1414.

Keith SJ. (2001). Evaluating characteristics of patient selection and dropout rates. *J. Clin. Psych.* 62 Suppl 9:11–14; discussion 15–16.

Maschio G; Oldrizzi L. (2000). Dietary therapy in chronic renal failure. (A comedy of errors). *J. Nephrol.* 13 Suppl 3:S1–6.

Organization, review, and administration of cooperative studies (Greenberg Report): a report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967. (1988). *Control Clin. Trials.* June 9(2): 137–148.

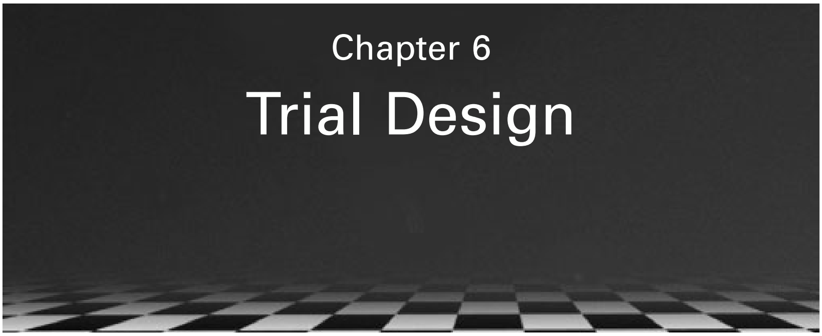
Weiss RB; Rifkin RM; Stewart FM; Theriault RL; Williams LA; Herman AA; Beveridge RA. (2000). High-dose chemotherapy for high-risk primary breast cancer: An on-site review of the Bezwoda study. *Lancet.* 355:999–1003.

Online Guidelines

http://www.ifpma.org/ich5.html

Chapter 6

Trial Design



Anyone who spends any time in a schoolroom, as a parent or as a child, becomes aware of the vast differences among individuals. My most distinct memories are of how large the girls were in the third grade (ever been beaten up by a girl?) and the trepidation I felt on the playground whenever we chose teams (not right field again!). Much later, in my college days, I was to discover there were many individuals capable of devouring larger quantities of alcohol than I without noticeable effect. And a few, very few others, whom I could drink under the table.

Whether or not you imbibe, I'm sure you've had the opportunity to observe the effects of alcohol on other persons. Some individuals take a single drink and their nose turns red. Others can't seem to take just one drink.

The majority of effort in *experimental design* is devoted to finding ways in which this variation from individual to individual won't swamp or mask the variation that results from differences in treatment. These same design techniques apply to the variation in result that stems from the physician who treats one individual being more knowledgeable, more experienced, more thorough, or simply more pleasant than the physician who treats another.

Statisticians have found three ways for coping with individual-to-individual and observer-to-observer variation:

1. ***Controlling*. Making the environment for the study—the patients, the manner in which the treatment is administered, the manner in which the observations are obtained, the apparatus used to make the measurements, and the criteria for interpretation—as uniform and homogeneous as possible.**
2. ***Blocking*. Stratifying the patient population into subgroups based on such factors as age, sex, race, and the severity of the condition**

and restricting comparisons to individuals who belong to the same subgroup.

3. ***Randomizing*. Randomly assigning patients to treatment within each subgroup so that the innumerable factors that can neither be controlled nor observed directly are as likely to influence the outcome of one treatment as another.[13]**

BASELINE MEASUREMENTS

In light of the preceding discussion, it is easy to see that baseline measurements offer two opportunities for reducing person-to-person variation.

First, some components of the baseline measurements such as demographics and risk factors can be used for forming subgroups or strata for analysis.

Second, obtaining a baseline measurement allows us to use each individual as his own control. Without a baseline measurement, we would be forced for our comparisons to depend on the final reading of the primary response variable alone.

Let's suppose this response variable is blood pressure. It might be that an untreated individual has a final diastolic reading of 90 mm Hg while an individual treated with our new product has a reading of 95 mm. Doesn't look good for our new product. But what if I told you the first individual had a baseline reading of 100 mm Hg, while the second had a baseline of 120 mm Hg. Comparing the changes that take place as a result of treatment, rather than just the final values, reveals in this hypothetical example that the untreated individual had a change of 10 mm, while the individual treated with our product experienced a far greater drop of 25 mm Hg.

The initial values of the primary and secondary response variables should always be included in the baseline measurements. Other essential baseline measurements include any demographic, risk factor, or baseline reading (laboratory values, ECG or EEG readings) that can be used to group the subjects of our investigation into strata and reduce the individual-to-individual variation.

CONTROLLED RANDOMIZED CLINICAL TRIALS

The trial design we will be most concerned with here is that of the long-term controlled randomized clinical trial. By *controlled* randomized clinical trial we mean a comparison of at least two treatment regimens one of which is termed a control.

---

[13] See, for example, Moore et al. (1998) and Chapter 8 of Good (2001).

Generally, though not always, as many patients will be assigned at random to the control regimen as are assigned to the experimental one. This sounds expensive and is. You're guaranteed to double your costs because you have to examine twice as many patients as you would if you tested the experimental regimen alone. The use of controls also may sound unnecessary. Your intervention works, doesn't it!

But the foreseen happens. You get the flu. You get a headache or the runs. You have a series of colds that blend one into the other until you can't remember the last time you were well. So you blame your silicon implants. Or you resolve to stop eating so much sugar. Or, if you're part of a clinical trial, you stop taking the drug.

It's then that as the sponsor of the trials you're grateful you included controls. Because when you examine the data you learn that as many of the control patients came down with the flu as those who were on the active drug. And those women without implants had exactly the same incidence of colds and headaches as those who had them.

Two types of controls exist: passive (negative) and active (positive):

A negative control or placebo in a drug trial may consist of innocuous filler, though the preparation itself should be matched in size, color, texture, and taste to that of the active preparation. A negative control would be your only option with disease conditions for which there is no existing "cure."

More often, there exists some standard remedy, such as aspirin for use as an anti-inflammatory, or metoprolol for use in alleviating hypertension. In such cases you would want to demonstrate that your preparation or device is equivalent or superior to the existing standard by administering this active preparation to the patients in your control group. Barbui et al. (2000) and Djulbegovic et al. (2000) recommend that an active control always be employed. Barbui et al. (2000) insist that to protect the patient, *only* an active control should be employed. Depending on your requirements and those of the regulatory agencies, one or both types of control may be needed.

Let's reflect on the consequences of not using controls. Who knows (or will admit) what executive or executive committee at Dow Corning first decided it wasn't necessary to do experimental studies on silicon implants because such studies weren't mandated by government regulations? It's terrifying to realize the first epidemiological study of whether breast implants actually increase the risk of certain diseases and symptoms wasn't submitted for publication until 1994, whereas the first modern implants (Dow Corning's Silastic mammary prosthesis) were placed in 1962.[14]

It's terrifying because the first successful lawsuit in 1984 resulted in a jury award of $2 million! Award after award followed with the largest ever, more than $7 million, going to a woman whose symptoms had begun even before she received the implants.[15] Today, the data from the controlled randomized trials are finally in. The verdict—silicon implants have no adverse effects on the recipient. Now, tell this to the stockholders of the bankrupt Dow Corning.

Randomized Trials

By randomized trial, we mean one where the assignment of a patient to a treatment regimen is not made by the physician but is the result of the application of a chance device, a coin toss, a throw of a die, or, these days, the computer program your statistician uses to produce a series of random numbers.

It may seem odd to circumvent the wisdom and experience of a trained physician in this way. Recall that the reason we are conducting trials is that we cannot yet state with assurance and government approval that one treatment regimen is better than another. Until our trials are completed and the data analyzed, there is no rational basis for other than a random assignment.

Warning. An investigator who has strong feelings for or against a particular regimen may not be an appropriate choice to work with you on a clinical trial. (See sidebar and Chapter 9.)

Blocked Randomization

Randomization means assigning treatment via a chance device. It does not mean give the first patient the active treatment, the second the control, and so forth. A weakness with this latter approach which you may have experienced yourself on the occasional visit to Las Vegas, Atlantic City, or Monte Carlo, is that sometimes red comes up seven times in a row or you experience an equally long streak with the dice.

In the long run everything may even out, but in the short run a preponderance of the subjects could get the active treatment, or in a multiple-site trial, one of the sites might have only control subjects

---

[14] According to Marcia Angell (1996), the recipient of the original implants still has them and has no complaints.

[15] *Hopkins v. Dow Corning Corp*, 33 F.3d 1116 (9th Cir. 1994).

BIAS

Do we really need to assign treatment to patients at random?

In the very first set of clinical data that was brought to me for statistical analysis, a young surgeon described the problems he was having with his chief of surgery. "I've developed a new method for giving arteriograms which I feel can cut down on the necessity for repeated amputations. But my chief will only let me try out my technique on patients he feels are hopeless. Will this affect my results?" It would and it did. Patients examined by the new method had a very poor recovery rate. But then, the only patients who'd been examined by the new method were those with a poor prognosis. The young surgeon realized he would not be able to test his theory until he was able to assign patients to treatment at random.

Not incidentally, it took us three more tries until we got this particular experiment right. In our next attempt, the chief of surgery—Mark Craig of St. Eligius in Boston—announced he would do the "random" assignments. He finally was persuaded to let me make the assignment using a table of random numbers. But then he announced that he, and not the younger surgeon, would perform the operations on the patients examined by the traditional method to "make sure they were done right." Of course, this turned a comparison of surgical methodologies into a comparison of surgeons and intent.

In the end, we were able to create the ideal double-blind study. The young surgeon performed all the operations, but his chief determined the incision points after examining one or the other of the two types of arteriogram.

assigned to it. In the first instance, a month-long epidemic of influenza could confound the effects of the epidemic with that of the treatment. In the second, the aspects of treatment unique to that particular site could be confounded.

To prevent either from happening, it is common to use a *block method of randomization*. The treatments are assigned a block of 8 or 16 patients at a time, so that exactly half of the patients in each block receive the control treatment. The assignment to treatment within each block is still in random order. But runs of "bad" luck are less likely to affect the outcome of the trials.

Caution: Blocked randomization can introduce bias to the study. See Berger and Exner (1999).

Stratified Randomization

If you anticipate differences in the response to intervention of males and females, of smokers and nonsmokers, or on the basis of some other important cofactor, then you will want to randomize separately within each of the distinct groups. The rationale is exactly the same as discussed in the preceding section: to ensure that in each group more or less equal numbers receive each treatment.

With life-threatening conditions the necessary data for stratification should be collected from each patient at the same time that eligibility is determined. This will permit assignment to treatment to be made as soon as eligibility is verified.

SINGLE- AND DOUBLE-BLIND STUDIES

A placebo is a pill that looks and tastes like the real thing but has no active ingredients. It functions though the power of suggestion. The patient feels better solely because he thinks he ought to feel better. It will not be effective if the patient is aware it is a placebo. Nor is the patient likely to keep taking the drug on schedule if he or she is told that the pill she is taking morning and evening contains nothing of value. She is also less likely to report any improvement in her condition, feeling the doctor has done nothing for her. Vice versa, if a patient is informed she has the new treatment she may feel it necessary to "please the doctor" by reporting some diminishment in symptoms. These sorts of behavioral phenomena are precisely the reason why clinical trials *must* include a control.

A *double-blind* study in which neither the physician nor the patient knows which treatment is received is considered preferable to a *single-blind* study in which only the patient is kept in the dark (Ederer, 1975; Chalmers et al., 1983; Vickers et al., 1997).

Even if a physician has no strong feelings one way or the other concerning a treatment, she may tend to be less conscientious about examining patients she knows belong to the control group. She may have other unconscious feelings that influence her work with the patients. For this reason one should also try to reduce or minimize contact between those members of your staff who are monitoring the outcome of the trials and those who have direct contact with the physician and her staff.

BREAKING THE CODE

An extreme example of how easy it can be to break the treatment code comes from a friend of mine who teaches at a medical school. He showed me a telegram he'd received from the company he was helping to conduct a clinical trial. They'd asked him to run an additional series of tests on half a dozen of the patients he'd been treating, including a PSA level. It didn't take a rocket scientist or my friend long to figure out that these had to be the patients who'd been given the drug under investigation. Not only had the trial sponsor broken the code by singling out some but not all of the patients for additional study, but they'd deliberately weighted the trials against their own product by failing to obtain an equal amount of adverse event data from the control population.

It is relatively easy (though occasionally challenging) to keep the patient from knowing which treatment she received. A near exception concerned trials of an early cholesterol-reducing agent which had the consistency and taste of sand; the only solution was to make the control preparation equally nauseous. Not unexpectedly, both treatment and control groups experienced large numbers of dropouts; few patients actually completed the trials.

Keeping the physician in the dark can be virtually impossible in most device studies, particularly if it is the physician who performs the initial implantation. With drugs, many physicians guess which patients are taking the active treatment (though not always correctly), and this knowledge may color their interpretation of adverse events.

A twofold solution recommends itself: First, whenever possible use an active control. A new anti-inflammatory should be compared to aspirin rather than placebo. Second, utilize two physicians per patient, one to administer the intervention and examine the patient,

INTENT TO TREAT

An obvious problem with a double-blind study is that it appears to rob the physician—the one closest to the patient—of any opportunity to adjust or alter the medication in accord with the needs of the patient. Thus many protocols provide for the physician to make an alteration when it is clearly in the patient's best interest.

Two policies preserve the integrity of the study even when such modification is permitted: First, the physician is not permitted to break the treatment code, lest she be tempted to extrapolate from the patient at hand to all those who received the same treatment. Second, the results from the patient whose treatment was modified continue to be analyzed as if that patient had remained part of the group to which he was originally assigned. Such assignment is termed "intent to treat" and should be specified as part of the original protocol.

As always, Bumbling Devices and Pharmaceutical carried the concept of "intent to treat" to an unwarranted extreme. At the onset of a single-blind study comparing two surgical procedures, a number of investigators performed the surgery first and only then looked at their instructions to see which modality ought to have been adopted. The result was a number of clear-cut protocol violations. (I place the blame for these violations not on the physicians but on the trial's sponsor for an inadequate training program.) Bumbling compounded their offenses and ensured total chaos by describing their study as "intent to treat" and reporting their results as if each patient had actually received the treatment she'd been assigned originally. One can only speculate as to the kind of penalties the regulatory agency ultimately imposed.

the second to observe and inspect collateral readings, such as angiograms, laboratory findings and X rays that might reveal the treatment.

Exceptions to the Rule

Are there exceptions to the rule? The regulatory agencies in some countries will permit variations from the fully randomized controlled long-term trial for certain highly politicized diseases (AIDS is a current example). But by going forward with such trials you run the risk that the results you obtain will be spurious and that after-market findings will fail to sustain or even contradict those obtained during the trials.

If you can't convince your boss of the risks that the failure to use controls may entail, may I recommend a gift of Marcia Angell's 1996 book on the saga of silicon breast implants.

SAMPLE SIZE

Determining the optimal sample size is neither as complex as outlined in statistics textbooks—for there is ample commercially available computer software programmed to remember the steps—or as simple—for the textbooks tend to omit too many real-world details. Seven steps are involved, and separately each step is quite easy to understand:

1. **Determine which formula to use.**
2. **Collect data to estimate precision.**
3. **Set bounds on the Type I and Type II errors.**
4. **Let the software make the initial calculation.**
5. **Determine the ratio of the smallest subsample to the sample as a whole.**
6. **Determine the numbers of dropouts, withdrawals, and noncompliant patients.**
7. **Correct the calculations.**

Which Formula?

We can expect to collect three types of data, each entailing a different method of sample determination:

1. **Dichotomous data such as yes/no, stenosis greater or less than or equal to 50%, and got better or got worse.**
2. **Categorical data (sometimes ordered and sometimes not) as we would see, for example, in a table of adverse events by type against treatment.**

COLLECT EXACT VALUES

At the beginning of a long-term study of buying patterns in New South Wales, it was decided to group the incomes of survey subjects into categories, under $20,000, $20,000 to $30,000, and so forth. Six years of steady inflation later, the organizers of the study realized that all the categories had to be adjusted. An income of $21,000 at the start of the study would only purchase $18,000 worth of goods and housing at the end. The problem was that those surveyed toward the end had filled out forms with exactly the same income categories. Had income been tabulated to the nearest dollar, it would have been easy enough to correct for increases in the cost of living and convert all responses to the same scale. But they hadn't. A precise and costly survey was now a matter of guesswork.

Moral. Collect exact values whenever possible. Worry about grouping them in categories later.

3. **Data such as laboratory values and blood pressure that is measured on a continuous metric scale.**

We should also distinguish "time-till-event" data (time till recovery, time till first reoccurrence) which, though metric, requires somewhat different methods.

From the point of view of reducing sample size, it is always better to work with metric data than categorical variables, and to work with multiple categories as opposed to just two. Even if you decide later to group categories, by income brackets, for example, it is always better to collect the data on a continuous scale (see sidebar). Some fine-tuning may still be necessary to determine which formula to use, but that's what statisticians are for.

Precision of Estimates

To determine the precision of our estimates for dichotomous and categorical data, we need to know the anticipated proportions in each category. For example, if our primary end point were binary restenosis we would need to know the expected restenosis rate. We would need such estimates for both control and treated populations. Usually we can collect data to estimate the former, but more often than not, we need to guesstimate the latter. One approach to guesstimating is to take the worst case, equal proportions of 50% in each treatment group, or, if there are multiple categories, to assume the data will be split evenly among the categories.

For metric data (other than time-till-event) it is common to assume a normal distribution, and to use the standard deviation of the variable (if known, as it frequently is for laboratory values) in calculating

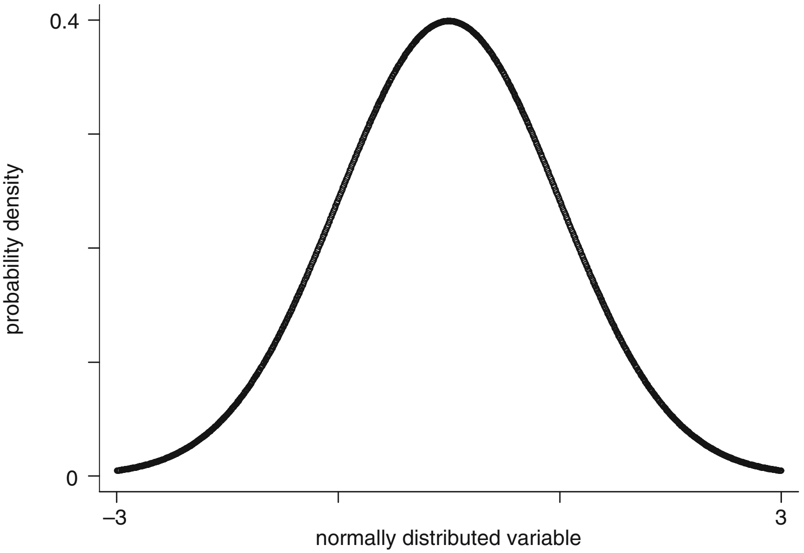


FIGURE 6.1 Bell-Shaped Symmetric Curve of a Normal Distribution.

the required sample size. You can see from Figure 6.1 that a normal distribution is symmetric about its mean and has a bell shape. Unfortunately, the distributions of many observations are far from symmetric (this is invariably the case with time-to-event data) and, more often than not, correspond to a mixture of populations—male and female, sensitive and less sensitive whose distribution resembles that of Figure 6.2.

Too often a normal distribution is used to estimate the necessary sample size, regardless of whether or not it is appropriate. If the data are unlikely to fall into a bell-shaped distribution, a bootstrap should be used to obtain the necessary estimate of sample size.[16]

For time-till-event data, an exponential distribution or one of the chi-square distributions may be used to calculate the required sample size. See also Therneau and Grambsch (2000, p. 61ff).

Bounding Type I and Type II Errors. A *Type I error* is made whenever one rejects a true hypothesis in favor of a false alternative. The probability of making a Type I error is called the *significance level.* A *Type II error* is made whenever one fails to reject a false hypothesis.

---

[16] See, for example, Manly (1992) and Tubridy et al. (1998). Details of the bootstrap method are given in Chapter 15.

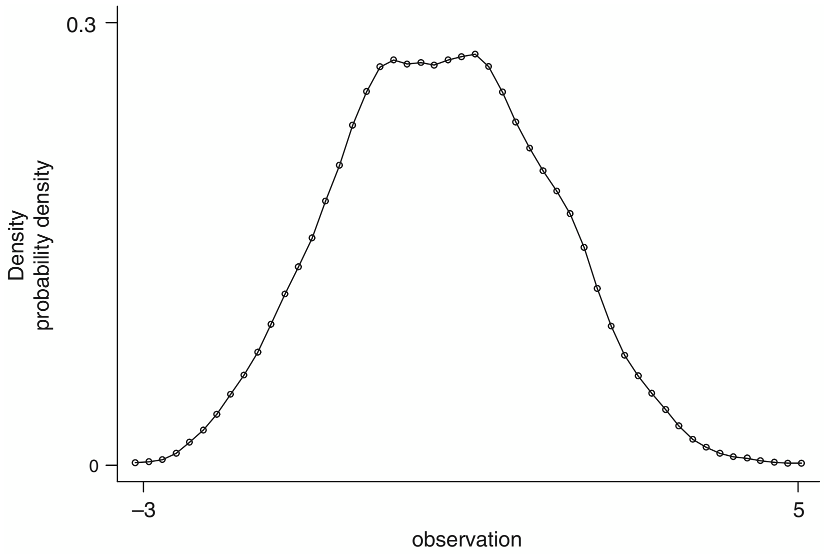


FIGURE 6.2 Mixture of Two Normal Distributions.

TABLE 6.1 Decision Making under Uncertainty

The Facts	Our Decision No difference	 Drug is better
No difference		Type I error Manufacturer wastes money developing ineffective drug
Drug is better	Type II error Manufacturer misses opportunity for profit Public denied access to effective treatment	

(See Table 6.1). When conducting a clinical trial, one can reduce but never eliminate these two errors.

Type I and Type II errors are interrelated. If we decrease the Type I error by using a statistical test that is significant at the 1% rather than the 5% level, for example, then unless we increase the sample size, our test will be less powerful and we are more likely to make a Type II error. The Type I and Type II errors, the difference between the treatment effect expected under the hypothesis and that expected under the alternative, and the sample size needed to bound the Type I and Type II errors are all interrelated.

To detect smaller treatment effects, we need more observations. We also will need more observations if we want fewer Type I or Type II errors. Consequently, in order to specify the required sample size, we first need to decide what size effect is really of interest to us, and what levels of Type I and Type II error can be tolerated. Here is an example:

Suppose that in 20% of patients, an untreated headache goes away within the hour. Do you want to demonstrate that your new headache remedy is successful in 21% of the cases? in 30%? in 50%? The less effective your new remedy, the larger the sample you will need to demonstrate a statistically significant improvement over the old.

You confront the same issues if your focus is on adverse events. The more rare the adverse event, the larger will be the sample you need to demonstrate a reduction.

The maximum allowable Type I error is normally specified by the regulatory agency. For establishing a primary response, 5% is customary. The Type II error or, equivalently, the power,[17] is under our control.

How certain do you want to be of detecting the improvement your treatment offers? Fifty percent of the time? Hardly adequate, not after you've spent several million dollars conducting the trials. Ninety percent? Ninety-five? Of course, you'd like to establish there is a difference 100% of the time, but unless your remedy works in 100% of cases, it can't be done, at least not without an infinite number of patients.

In the end, the bound on Type II error you and your statistician arrive at may have to represent a compromise between how much you are willing to spend on the trials and how reluctant you are to let a promising remedy slip through your hands.

Equivalence. The preceding discussion is based on the premise that you want to show that your treatment represents an improvement over the old one. If, instead, you want to demonstrate equivalence,[18] then you may want to keep the sample size as small as possible. For as the sample grows larger and larger, you are guaranteed to reject the hypothesis of equivalence no matter how small the actual difference between the effects of the two treatments may be.

---

[17] The *power* of a test is defined as the compliment of the probability of making a Type II error, that is, the probability of correctly rejecting a false hypothesis and accepting the alternative. The more powerful the test, the smaller will be the Type II error.

[18] When we say two treatments are equivalent, we don't really mean they have identical effects, merely that they are sufficiently close in effect that physiologically they cannot be told apart.

Whether you are testing that two treatments are equivalent or that one is superior to the other, a sample size that is adequate for establishing a treatment's *efficacy* may not be adequate for establishing its *safety*. More and more often, regulatory agencies are imposing a bound on the Type II errors allowable when reporting adverse events. The more rare and more severe the event, the larger the sample required to eliminate fears of its occurrence. In the end, this bound rather than your own needs may determine your sample size.

Software. We've listed some of the computer software that can be used for sample size calculations such as nQuery, Pass 2000, and StatXact in an Appendix. For additional details on the methods of calculation, see Shuster (1993).

Subsamples. The number that the software's calculations yield may appear rather smaller than you expected. That's the good news. The bad news is that this number may represent only a small fraction of the sample you'll need to consider when all cofactors are taken into consideration. The worst case occurs when you expect to find differences in the magnitude (and perhaps even the direction, plus or minus) of the treatment effect depending on whether the subject is male or female, a smoker or nonsmoker, has multiple risk factors or none at all, and so forth. When you are almost certain such *interactions* don't exist (you can never be absolutely sure), then rather than double the total sample size, you may only need to increase it by one or two additional observations for each cofactor.

In the worst-case scenario you'll need to double the sample size you calculated initially to include both males and females, double it again to include smokers and nonsmokers, double it a third time to include treated subjects as well controls, and so on; the final result may be undesirably large.

Loss Adjustment. Alas, even this result is not final, for it represents only those subjects who will complete the study. You can expect to loose many patients along the way.

The greatest losses, sometimes exceeding 90%, occur during the initial screen. The cost to you of these ineligibles is thus a minimum; still, they have to be accounted for. Unfortunately, the clinical literature provides no consistent guidelines for determining the exact percentage of those interviewed who will go on to actually participate in the study.

You could rely on your past experience or that of your colleagues. Don't hesitate to do what you can to improve the numbers. In Chapter 9, we discuss several methods for increasing the percentage you enroll—but you will need to multiply the numbers you came up with originally to correct for ineligibles.

The second set of losses occurs once the trials are under way as a result of dropouts and noncompliance. Patients don't remain on their medication, or begin to self-dose with some unrelated drug. Patients don't show up for appointments. Or patients may drop out of the trials altogether. In Chapters 9 and 12 we discuss methods you can employ to reduce these losses. Again, you must factor in these yet-to-occur losses before settling on a final figure for the size of the initial sample.

NUMBER OF TREATMENT SITES

If all the information you needed could be collected at a single site, the result would be an immediate reduction in patient-to-patient variation with an accompanying decrease in the necessary sample size. But that's not going to happen, particularly if the disease condition you hope to treat is a relatively rare one. Even if it could happen, you might want to use multiple sites if you feel it would reduce the total time required to complete the trials. And multiple sites can mean a more diversified patient population.

Either way you need to have some advance notion of the number of patients you might hope to treat at each site, which means you have to have some notion of the prevalence of the disease condition. If you come up with less than six patients per site, then you will need to increase the duration of your study in order to enroll a sufficient number of eligible patients (see also Chapter 9). Once you have this number in hand, you can divide the sample size by it to determine the number of physicians you will need to recruit.

PRACTICAL STEPS YOU CAN TAKE TO REDUCE LOSSES

- Target recruiting efforts at those both eligible and likely to participate. See Chapter 9.
- Review and, if possible, loosen eligibility requirements.
- Select appropriate participants, meaning those most likely to remain in compliance
- Establish measures to increase compliance. See Chapter 5.

ALTERNATE DESIGNS

KISS is the operative phrase in the design of the large-scale long-term randomized controlled clinical trials that are our chief concern in this guide. We advise you to resist all attempts to measure redundant variables ("but surely as long as the patient is in my office you won't mind if I perform one or two tests of my own") or complicate the design. On the other hand, short-term clinical trials of more limited scope whose objective is to determine the maximum tolerable dose or to establish efficacy can often benefit from the use of more complex designs such as a crossover or a fractional factorial.

In a *crossover design*, each patient receives all treatments in order, treatment A followed by treatment B, or treatment B followed by treatment A (or, if there are more alternatives, A followed by B followed by C). Thus each patient serves as her own control, reducing the individual-to-individual variance to an absolute minimum.

In a *fractional factorial design*, best employed when there are adjunct treatments and/or multiple cofactors, only some and not all treatment combinations are tested. Sophisticated statistical methods are used during the analysis phase to compensate for the missing data.

The advantage both of these design types offer is that they markedly reduce the total number of patients required for the trial. Their disadvantage, again in both cases, is that their validity rests on certain key assumptions that are seldom realized in practice.

To use a crossover design, one has to assume that neither treatment has a residual effect, that using B after A has exactly the same effect on a patient as if A had never been used.[19] In particular, one has to assume that trace quantities of A or B and their metabolic by-products do not linger in the body after treatment with A or B is ended. If crossover trials are contemplated, a pharmacokineticist is an essential addition to the design team.

To maintain the validity of a fractional factorial design, one has to be able to assume that the effect of treatment A is the same at all levels of the cofactor and in all subgroups. Again, these assumptions are seldom realized in practice and represent major drawbacks for the methodology.

But the main objection to these designs is that full-scale long-term clinical trials have not one but two purposes: to demonstrate *both* efficacy and safety. A sample size that might be adequate for demonstrating the one may be far too small to establish the other. The chief advantage of crossover and fractional factorial designs—reduction in sample size—is lost while their disadvantages remain.

---

[19] Or that the two treatments have the same residual effects, an even more unlikely prospect.

A third type of study, occasionally used to demonstrate efficacy, employs *case controls*. The data for these controls are obtained by referencing historical databases and attempting to find patients whose profiles (demographics, risk factors, laboratory values) align as closely as possible with those of patients who received the investigational intervention. As the allocation of patients to treatment was not made at random, nor were the treatments of control and experimental subjects contemporaneous, this type of design is not appropriate for full-scale clinical trials. They can be useful in demonstrating to the regulatory agency the validity of going forward with large-scale clinical trials (see Chapter 8).

If death is a possible outcome for an untreated or inadequately treated patient as it is, for example, with AIDS, you may want to consider the use of *response adaptive randomization* in which the majority of new patients are assigned to the currently most successful treatment. If the "success" was temporary or merely a chance event, then the proportions will gradually even out again or perhaps go the other way. But if further trials sustain the advantages of one treatment over another, then a greater and greater proportion of patients will be assigned to the preferable treatment and the number of deaths during the trials will be kept to a minimum.

The analysis of such trials is complicated, but it is well understood and thoroughly documented; see, for example, Yao and Wei (1996). The chief drawback is the lack of commercially available software with which to perform the analysis.

TAKING COST INTO CONSIDERATION

The chief controllable factors affecting the cost of clinical trials are:

- **Choice of end points**
- **Methods of measurement**
- **Data entry**
- **Eligibility criteria**
- **Patient recruiting methods**
- **Physician reimbursement**
- **Duration of the trials**

Profit considerations should be taken into account when making decisions about the design of randomized controlled trials. For example, more precise measurements are generally more costly but their use can reduce the number of patients that are required. A cost analysis of all alternatives should be made before the final of choice of end points is made.

Sample sizes need not be balanced. A design that assigns more patients to the less costly treatment group can be more cost effective. See, for example, Torgerson and Campbell (1997).

Computer-aided direct data entry will result in a substantial reduction in costs and, along with computer-aided NDAs, will increase profits by allowing you to bring a product to market sooner; see Chapters 8 and 10. Chapter 9 contains a number of suggestions for making patient recruiting efforts more cost effective.

Economic models can be used to determine a portfolio of studies that maximizes the expected return on a given development or trial budget. See, for example, Backhouse (1998), Cavan (1995), and Claxton and Posnett (1996).

FOR FURTHER INFORMATION

Backhouse ME. (1998). An investment appraisal approach to clinical trial design. *Health Econ*. 7:605–619.

Barbui C; Violante A; Garattini S. (2000). Does placebo help establish equivalence in trials of new antidepressants? *Eur Psychiat*. 15:268–273.

Berger VW; Exner DV. (1999). Detecting selection bias in randomized clinical trials. *Control. Clin. Trials* 20:319–327.

Berlin JA; Ness RB. (1996). Randomized clinical-trials in the presence of diagnostic uncertainty—Implications for measures of efficacy and sample-size. *Control. Clin. Trials* 17:191–200.

Cavan BN. (1995). Improving clinical trials cost management in biotech companies. *Biotechnol*. (NY) 13:226–228.

Chalmers TC; Celano P; Sacks HS; Smith H. (1983). Bias in treatment assignment in controlled clinical trials. *N. Engl. J. Med.* 309:1358–1361.

Claxton K; Posnett J. (1996). An economic approach to clinical trial design and research priority-setting. *Health Econ*. 5:513–524.

Djulbegovic B; Lacevic M; Cantor A; Fields KK; Bennett CL; Adams JR; Kuderer NM; Lyman GH. (2000). The uncertainty principle and industry-sponsored research. *Lancet*. 356:635–638.

Ederer F. (1975). Why do we need controls? Why do we need to randomize? *Am. J. Ophthalmol*. 76:758–762.

Elwood JM. (1998). *Critical Appraisal of Epidemiological Studies and Clinical Trials*, 2nd ed. New York: Oxford University Press.

ICH. (1998). *E 9 Statistical Principles for Clinical Trials*. *Fed. Reg*. 63:49583. *http://www.fda.gov/cder/guidance/91698.pdf*.

ICH. (2001). *E 10 Choice of Control Group and Related Issues in Clinical Trials*. *http://www.fda.gov/cder/guidance/4155fnl.htm*.

Jones B; Kenward MG. (1989). *Design and Analysis of Crossover Trials*. New York: Chapman and Hall.

Manly BFJ. (1992). Bootstrapping for determining sample sizes in biological studies. *J. Exp. Mar. Biol. Ecol.* 158:189–196.

Matthews JNS. (2001). *An Introduction to Randomized Controlled Clinical Trials.* Oxford: Arnold Press.

Moye LA. (1998). *P*-value interpretation and alpha allocation in clinical trials. *Ann Epidemiol*. 8:351–357.

Piantadosi S. (1997). *Clinical Trials: A Methodologic Perspective*. New York: Wiley.

Shuster JJ. (1993). *Practical Handbook of Sample Size Guidelines for Clinical Trials*. Boca Raton, FL: CRC Press.

Simon R. (1999). Bayesian design and analysis of active control clinical trials. *Biometrics* 55:484–487.

Thall PF; Cheng SC. (2001). Optimal two-stage designs for clinical trials based on safety and efficacy. *Stat Med*. 20:1023–1032.

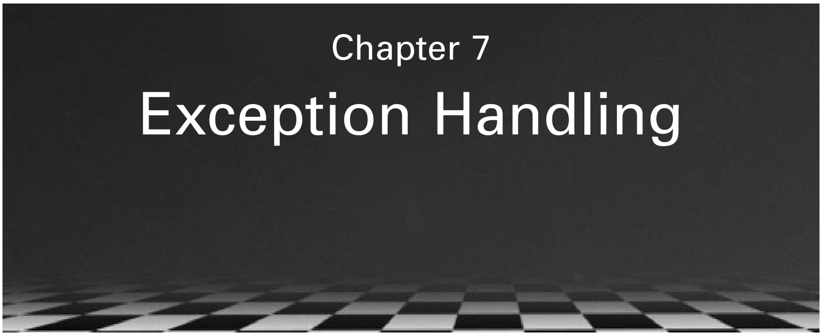
Therneau TM; Grambsch PM. (2000). *Modeling Survival Data.* New York: Springer.

Torgerson D; Campbell M. (1997). Unequal randomisation can improve the economic efficiency of clinical trials. *J. Health Serv. Res. Pol*. 2:81–85.

Tubridy N; Ader HJ; Barkhof F; Thompson AJ; Miller DH. (1998). Exploratory treatment trials in multiple sclerosis using MRI: Sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups. *J Neurol Neurosurg Psychiat.* 64:50–55

Vickers A; Cassileth B; Ernst E; and et al. (1997). How should we research unconventional therapies? *Int. J. Tech. Assess. Health Care* 13:111–121.

Yao Q; Wei LJ. (1996). Play the winner for phase II/III clinical trials. *Stat Med*. 15:2413–2423; discussion 2455–2458.



Chapter 7

Exception Handling

THIS CHAPTER IS DEVOTED TO PLANNING for the innumerable petty but essential details—missed appointments, patient complaints, and protocol deviations—that inevitably arise in an extensive and lengthy series of clinical trials. We also consider certain more serious matters such as a high frequency of adverse events that may result in early termination of your study.

PATIENT RELATED

Missed Doses

Phone calls to investigators from patients who have missed a scheduled dose are common. A uniform policy on missed doses should be incorporated in both patient and investigator instructions.

Missed Appointments

Missed appointments are commonplace also, with noncompliant patients being particular offenders. Establish a policy of prior notification by the investigator's office (perhaps a card in the mail a week before the visit, and a telephone call the day before). Once again, having a sponsor-paid coordinator at each site helps ensure that your policies are adhered to.

Patients, particularly those whose health has improved or who dislike the treatment, cannot be counted on to reschedule on their own. Have site coordinators follow up immediately by telephone should the patient not appear at the scheduled time.

Suppose (and one should always suppose the worst case as it is inevitable) that the patient fails to appear for the one-month follow-up exam but does appear at some time prior to the two-month follow-up. If the patient shows up at five weeks, would this be close enough in time to count as the one-month follow-up? If the patient appears for the first time at seven weeks, would you mark the one-month follow-up as missing and record this exam as the two-month follow-up?

How will you treat patients who show up at other intermediate times? You and your design team need to formulate a consistent policy that will be adhered to throughout the study.

Noncompliance

Noncompliance of patients with the treatment regimen has four chief sources:

1. **Ambiguous directions**
2. **Noncooperative or frightened patients**
3. **Unreported use of concurrent medications**
4. **Unpalatable or otherwise undesirable therapy**

The first two of these can be dealt with by careful attention to detail during the preparation of patient instructions and in the training of personnel who will have direct contact with patients. To deal with the third, questions on concurrent medications, both prescribed and self-administered, should be made part of each follow-up examination. The last *should* have been detected and forestalled during preliminary trials. But if the reality is that patients are reducing or skipping doses because of taste, stomach upsets, or other adverse events, the results will be effected. Careful questioning during follow-up examinations is essential. (See Chapter 14.)

Adverse Reactions

Physicians are acutely aware of the rare but inevitable instances in which a patient has an immediate adverse reaction to treatment or to the collection and diagnostic procedures associated with treatment. Similarly any surgical intervention may be accompanied by undesirable events not directly related to the procedure under investigation. You will need to list all such possible reactions and prepare written procedures for dealing with them. This list will become part of your written submission to the regulatory agency.

Reporting Adverse Events

You will require a separate form for recording each adverse event that occurs during the study and for (possible) reporting to the

FORMULATE AN ADVERSE EVENTS POLICY

The policies of the Committee for the Protection of Human Subjects (CPHS) at the Naval Health Research Center in San Diego provide an excellent starting point. Their policy, reproduced below, defines an adverse event, describes who should be notified and under what circumstances, and also provides the name of individuals who should be contacted if additional guidance is necessary.

An adverse event is "any incident, accident, untoward drug reaction, appearance of disease or injury which may have occurred, or which could reasonably have occurred, as a result of using research volunteers in research.

When an adverse event occurs, the Principal Investigator (PI) should notify the Commanding Officer by e-mail or fax within 24 hours of the incident, and an informational copy of the same e-mail should be sent to the CPHS Chair. The PI should follow this up with a Memo to the CPHS Chair that follows the format shown under "Adverse Event Reporting" on the NHRC Intranet.

Subsequently, the CPHS will convene to consider the report and to decide if changes to the protocol are needed. If you have any questions about this procedure, please ask. . . ."

regulatory agency. This form should provide for both anticipated events (nausea, headache) and unanticipated (other), for the trivial (nausea, headache) and the serious.

The form should note whether the event is continuing or preexisting and (in the investigator's opinion) to what degree it might be related to the intervention. Action(s) taken and their outcome should be noted along with links to any secondary sets of forms that may have been completed. The appropriate regulatory agency requirements should be consulted.

Investigators should be instructed to complete and transmit adverse event forms to you as soon as they become aware of the event. As always, computer-assisted data entry facilitates both completion and transmission of such forms.

When Do You Crack the Code?

On receipt of an adverse event form, it should be collated with the set of forms that have already been submitted and two questions addressed:

1. **Are most of the events taking place at a specific site or sites?**
2. **Is a particular event or pattern of events occurring with unusual frequency?**

If the majority of adverse events are occurring at a particular site or sites, the response of your CRMs should be appropriate to the several possibilities: If those sites are treating the majority of the patients, a high proportion of adverse events is to be expected and no further action is required. If those sites are the most conscientious in recording adverse events, the importance of tracking adverse event needs to be stressed with the site coordinators at the remaining sites. If these latter sites may be deviating from the protocol, a visit is warranted.

How you react to an unusually high frequency of adverse events will depend on the severity of the events and whether they were expected or unexpected. An external review panel whose primary concern is the safety of the treatment should review the data concerning the events. The members of this panel should not be regular employees of your company; their skills should mirror those of the members of your design team. Upon this panel's recommendation, the code may be broken and the data in hand subjected to a comprehensive statistical analysis.

Chapter 14 contains a further discussion of this important issue.

INVESTIGATOR RELATED

Lagging Recruitment

Enrollment should be monitored on a continuous basis. Fortunately there is a great deal of commercially available software to help you in this task (see the Appendix). Forecasting methods are described in Chapter 14.

Eligibility forms should be completed and transmitted to you on the same day that the patient is examined. If only a few sites have lagging enrollments, you are free to concentrate your efforts on those sites. If recruitment is an across-the-board problem, you have five alternatives:

1. **Increase the time allotted to complete the trials**
2. **Launch an intensive recruiting campaign (see Chapter 9)**
3. **Recruit additional study centers**
4. **Modify the eligibility requirements**
5. **Abandon the trials**

Prepare for the worst and have a backup plan ready.

Protocol Deviations

Potential protocol deviations include all of the following:

- **Enrolling ineligible patients**
- **Failing to ensure each patient has made informed consent**

- **Initiating an intervention other than the one assigned**
- **Altering the nature of the intervention without permission or notification to you**
- **Failing to record data in the manner specified or at the specified times**
- **Recording fraudulent data**

Preventive measures include:

- **Monitoring enrollment procedures**
- **Keeping the intervention and the measurements to be made straightforward and easy to follow**
- **Preparing a detailed, yet easy-to-follow procedures manual (see Chapter 8)**
- **Providing a comprehensive training program (see Chapter 10)**
- **Monitoring the data being collected (see Chapters 12 and 13)**

Site-Specific Problems

When lagging enrollments, ineligible patients, patient dropouts, delays in transmittal of forms, protocol deviations, excessive numbers of adverse events, or evidence of fraud can be traced to a few specific sites, then the first obvious step is a visit to the site by the clinical research monitor. In most instances, problems can be resolved through such visits. Perhaps additional training is required, including a detailed walk through the various intervention and recording procedures. Perhaps a visit by the medical monitor to the offending physician(s) may be warranted.

Should such friendly persuasion prove unavailing, you will need to have a plan in place. For minor offenses—too many ineligible patients, too many dropouts, delays in submitting forms—the solution is to do what you would do with any recalcitrant but momentarily essential employee—discontinue further enrollment at the site, spend whatever additional time is necessary to ensure compliance with those patients that are already enrolled, and, above all, pay only for correctly completed forms.

With more serious offenses, your choices are more limited. You will need to notify the regulatory agency of any deviations. The regulatory agency will require that you continue to provide treatment for and monitor the progress of those patients that are currently under the offending physician's care. Legal as well as ethical issues are involved. You don't want to be there.

Preventive measures are essential and include all of the following:

- **Care in recruitment of study physicians and laboratories (see Chapter 9)**

- **Drafting contracts with study physicians and laboratories that spell out the procedural requirements and the penalties for violating them**
- **Monitoring the data being collected and taking the earliest possible remedial action**

Closure

In the previous chapter we discussed the possibility of an unplanned closure dictated by a high frequency of adverse events. Seldom does an interim analysis reveal a clear-cut pattern: treatment bad but control good, or vice versa. More often, the results suggest that external factors are responsible for adverse events or that the events are affecting only a single subgroup such as those with specific risk factors or the most preexisting complications. In such an instance you may wish to stop enrolling any further members of that subgroup in the study. Only in the event that a single treatment arm appears to be deleterious for all subgroups should the intervention be discontinued and the patients assigned to another treatment arm or released to the care of their personal physician.

Note. If you discontinue treatment to all patients, you are obligated to notify the regulatory agency and to continue to monitor trial subjects until the scheduled time for termination is reached. Further discussion is in Chapter 14.

Intent to Treat

When an intent-to-treat regimen is adopted, the physician is free to modify or withdraw treatment if warranted by the patient's condition. To turn theory into practice, guidelines for modifications should be established in advance of the trials. Is the physician free to alter the dosage? Or to add an adjunctive therapy? Is he restricted to switching among the protocol alternatives or may he switch to any treatment he deems appropriate? Have your answers ready before the trials begin.

IS YOUR PLANNING COMPLETE? A CHECKLIST

Determined the following:

- **Study objectives**
- **Primary responses (efficacy)**
- **Secondary responses (safety)**
- **Baseline variables**
- **Study population**
- **Time line**
- **Closure**
- **Who will do the monitoring?**

Group observations by the individual (or lab) making the observations and the time of collection. Began recruiting for implementation team and study review panels.

The completed trial design should provide for:

Controls
Randomization
Blinding
Intent to treat
Sample size

And it should provide for exceptions. You should know who will monitor, who will respond to you, and how you will deal with the following:

Patient-related exceptions
Investigator-related exceptions
Unexpectedly high frequencies of adverse events
Serious adverse events
Protocol violations

Part Two

DO

Chapter 8

Documentation

THIS CHAPTER DESCRIBES the documentation you must make for the regulatory agency and should make for yourself and your coworkers.

To go ahead with the trials you will need to submit a *proposal* to the regulatory agency.

If you have followed our prescription so far, your design committee will already have prepared a *protocol*, and you will be in the process of designing *procedure manuals* for your investigators and the templates for a series of *interim reports* for use by your staff when monitoring the trials.

Once the trials are complete, you will need to submit one or more *final reports* to the regulatory agency. You also will need to submit one or more *interim reports* to them if you are compelled by circumstance to alter the nature of the trials or to terminate the trials prior to completion. Marketing may ask that you seek to publish your findings. An *AAR* (After Action Review), discussed in the final chapter of this text, is the essence of good management.

We cover the scope and contents of all these reports in what follows.

GUIDELINES

Two fundamental rules govern all your written communications with individuals outside your company:

1. **If you and the members of your staff can't follow a report, neither will outside reviewers. The result of sending out such a report will be both potential miscommunication and substantial delays.**
2. **Your submission is a contract. Do not commit to tasks you cannot honor or state facts you cannot support.**

STUDY PROTOCOL TABLE OF CONTENTS

- Sponsor Data
- Introduction: Background and Rationale
- Objectives
- Patient Selection
- Treatment Plan
- Outcome Measures and Evaluation
- Procedures
- Clinical Follow-Up
- Adverse Events
- Data Management, Monitoring, and Quality Control
- Early Withdrawal
- Statistical Methods
- Investigator Responsibilities
- Ethical and Regulatory Considerations
- Study Committees
- Appendixes
  - Bibliography
  - Sample Informed Consent Form
  - Schedule of Events

INITIAL SUBMISSION TO THE REGULATORY AGENCY

Not surprisingly, the form of the protocol or investigational plan to be submitted to the regulatory agency closely mirrors the design elements discussed in Chapters 5 and 6. Most of the headings on the accompanying model table of contents should be familiar to you. (The order, content, and titles used in this sample table may need to be varied depending on the requirements of the regulatory agency.)

Sponsor Data

Sponsor data should include the name, address, and telephone number of your company, the name and title of the chief investigator, medical monitor, medical safety officer, and the name, title, phone number, and email address of your regulatory agency liaison.

Justifying the Study

Your justification should be that originally presented to the executive committee, comprehensive yet summarized in no more than a page or two. Begin by stating the prevalence of the disease condition you propose to treat along with its effects. Here is an example: “Colorectal cancer is the second most common visceral malignancy in the United States. An estimated 156,000 new cases of the disease will occur. . . .”

The balance of your justification should briefly summarize previous work you and other investigators have done that would lead one to believe that the intervention you propose will be both safe and efficacious. Include data from any or all of the following sources:

- **Pharmacology and biochemical theory—describe the mechanism of action if known**
- **Animal experiments—include toxicology findings**
- **Anecdotal studies**
- **Case-control studies**
- **Short-term clinical studies including Phase I determinations of the maximum tolerated dose and Phase II studies of the minimum effective dose.**

You should also reference any previous full-scale clinical studies when you are proposing extensions of the subject population, modifications to the treatment regimen, or new indications for use of an already marketed intervention.

Brief descriptions of each study should be provided along with journal and text references where available.

Again, note that what distinguishes the full-scale clinical study from prior clinical trials is that it entails measures of both safety and efficacy, involves a predetermined dose or treatment regimen, is long-term, and includes sufficiently many patients that accurate estimates of the incidence of all but the rarest of side effects can be made.

Take pains to differentiate your proposal from prior work. For example, "The proposed study will include both men and women. It will involve more than 10 times the number of patients observed in any single previous study. The study period of one year is eight months in excess of any previous study period."

Remember, the object of your proposal is to convince even the most conservative and cautious reader that the investigation you propose is both prudent (because it is solidly grounded in prior efforts) and desirable (because the study's depth and breadth ensures the public will be fully protected once the study is complete).

Objectives

This section as well as each of the remaining sections of your proposal should be largely self-standing. While this will lead to substantial redundancy, it facilitates the review process and forestalls misunderstandings.

This section should begin with a focused restatement of what you outlined in your introduction and include brief definitions of the primary and secondary endpoints.

Patient Selection

Begin with a summary description of your intended population, for example, patients with histologically proven colon or rectal cancer at the time of surgery. Then provide comprehensive listings of the inclusion and exclusion criteria you will employ. (Reminder: Any limitations will also limit the scope of subsequent marketing of the intervention.)

Here is an example:

> ***Eligibility Criteria.*** **To be eligible for this study a patient must satisfy the following criteria:**
>
> - **Be between 30 and 80 years of age.**
> - **Have a physician's assessment of good general health with an expected survival of at least five years.**
> - **Have the ability and willingness to understand informed consent and to comply with study procedures.**
> - **Not be currently undergoing treatment in a chemoprevention trial.**
> - **Not be pregnant or nursing. Women of childbearing potential must be willing to use an effective method of birth control throughout the study.**
> - **Have histologically proven colon or rectal cancer."**
>
> ***Exclusion Criteria.*** **To be eligible for this study, a patient may not have any of the following diagnosed health conditions:**
>
> - **More than 100 polyps at the time of resection.**
> - **Active invasive malignancy, other than nonmelanoma skin cancer.**
> - **Cardiovascular disease—NYHA Class 3 or 4.**
> - **Immunosuppressive therapy within the six months preceding the intake appointment. Nonimmunosuppressive therapy steroid therapy does *not* necessitate exclusion from the study.**
> - **Clinically obvious narcotic and/or alcohol dependence within the six months preceding the intake appointment.**
> - **History of ulcerative colitis or Crohn's disease.**

TREATMENT PLAN

This section should include the following:

- **Time line for the trials (include a diagram)**
- **Brief description of the experimental design including the method of randomization, the extent of the blinding and the method(s) to be employed to ensure the blinding is sustained throughout the course of the trials**
- **Sample size**
- **Method of treatment assignment**
- **Rules governing early termination or modification of the protocol.**

OUTCOME MEASURES AND EVALUATION

This section incorporates a more detailed and precise discussion of the end points listed under the heading of objectives. Describe how and by whom each measurement will be taken and how and by whom each measurement will be evaluated.

PROCEDURES

This section includes brief descriptions of any invasive procedures as well as of specimen collection and handling. It should also include methods for handling any possible adverse reactions to procedures. For studies involving medications, the following must be included:

- **Dosage form (capsule, tablet, ointment)**
- **Route of administration (oral, intramuscular, intravenous)**
- **Frequency of administration**

For intravenous administration include the rate of administration and the concentration of the medication in the delivery medium.

Clinical Follow-Up

Provide tabular summaries of any required testing and any other follow-up procedures as in Table 8.1. In other words, expand on the time line covered under the treatment plan. Also you will need to describe:

- **Any concomitant medical therapy**
- **Procedures for dealing with withdrawals and noncompliant patients**
- **Extent of any further follow-up examinations after the termination of the study.**

TABLE 8.1 Schedule of Events

	Screen	Surgery	Two Weeks	One Month	Six Months	Eight Months	One Year
History	√						
Consent	√						
ECG	√	√			√		
WBC	√	√	√	√			
Creatinine	√	√					
Cholesterol	√						
CK&CKMB	√	√					
ACT		√					
Follow-up							
Angiogram	√	√	Phone	√	√	√	√

ADVERSE EVENTS

List the most likely adverse events along with how you plan to monitor and report on them. Here is an example from a study of the effects of aspirin.

> **A questionnaire will be completed for each patient at each follow-up visit. The questionnaire specifically addresses gastrointestinal symptoms (nausea, vomiting, heartburn, dyspepsia), abdominal pain, and bleeding.**
>
> **Patients will also be encouraged to report any side effects as they occur and will be provided with an information sheet containing the local coordinator's telephone number.**
>
> **A physical examination will be conducted at each follow-up visit and any new medical conditions will be recorded. Laboratory tests will be performed as the physician determines.**

DATA MANAGEMENT, MONITORING, QUALITY CONTROL

Briefly describe the use of computer-aided data entry. State that all data entry screens will be incorporated as part of the final submission. (See under E-submissions later in this chapter.)

STATISTICAL ANALYSIS

Describe the analytical measurement(s) to be made, the relevance to the protocol objectives, the populations to be analyzed, and the statistical methodology to be utilized. Be brief—a textbook is not required. Specify the analytical plan to be used for the protocol measurement(s). Include the criteria and procedures used to assess analytical results. Cite all relevant scientific literature supporting the use of the analytical method for the intended measurements.

Here is an example:

> **The difference in recurrence rates between the aspirin and the placebo groups will be tested using Fisher's Exact Test. A preliminary exact test will be performed to see if the data from all cites may be combined. (See Good, 2000.)**
>
> **Sample size was chosen so as to have 90% power to detect a 10% difference in recurrence rates with a test at the 5% significance level.**[20]

---

[20] The statistical terms were defined in Chapter 6.

FILL IN THE HOLES

Your protocol should not only describe *how* the data are to be analyzed but *which* data are to be used for the analysis. Take into account *all* contingencies. Here's an example from an (almost) successful study.

The appropriate angiogram to use for follow-up purposes is to be determined as follows:

1. If a patient had a target site revascularization after 14 days and prior to 7.5 months, the angiogram immediately preceding the surgery will be used.
2. If not, then the first follow-up angiogram within eight ± 0.5 months after the index procedure will be used.
3. If such an angiogram is not available, then use a follow-up angiogram performed at least four months after the index intervention.
4. Otherwise, if there is objective evidence of recurrent ischemia between 14 days and four months and repeat angiography during the same period demonstrates restenosis, that angiogram will be analyzed as the follow-up angiogram.
5. If such angiograms are not available, then an angiogram taken within 12 months will be used.

INVESTIGATOR RESPONSIBILITIES

This section should include explicit statements on what the investigator will *not* do:

- **Utilize the experimental intervention outside the scope of the study**
- **Undertake investigative procedures on any enrolled patient other than those specified in the protocol**
- **Publish the results of their experience with the intervention until the publication of the multiple-center results**

However, the investigator is obligated to:

- **Administer and obtain informed consent from all patients**
- **Administer the intervention and record and report measurements only as specified in the protocol**
- **Submit all data and supporting documents in a timely fashion**
- **Return any unused drugs or devices at the conclusion of the study**
- **Retain copies of all patient records as specified by the regulatory agency (and your own) requirements.**

This section should also be incorporated in the procedures manual you give to each physician.

IF YOU DON'T HAVE ANYTHING TO SAY . . .

The protocol is not a term paper, your objective not merely to cover a piece of paper with words, but to provide a detailed description of what you propose to do. KISS.

Consider the following example of what *not* to write taken in its entirety from an actual Bumbling proposal:

Secondary Response Variables and Multi-variable Modeling

Many secondary response variables may be evaluated. The measurable responses fall under two broad testing categories:

1. Subgroup hypotheses that utilize the primary hypothesis end points for potentially important patient strata, such as long stints, use without extended surgery, restenosis in women, diabetics, or the left descending artery
2. Secondary hypotheses that require separate end points for analysis (e.g., costs or vascular complications). The secondary questions will be confined to three categories: other definitions of restenosis, complications, or other.

ETHICAL AND REGULATORY CONSIDERATIONS

This section is required by some but not all regulatory agencies and its form and content will depend on which branch of the regulatory agency is responsible for supervising your trials. Analogous to the preceding section, it consists of a listing of your company's responsibilities with respect to the patients, the investigators, and the regulatory agency. It is a form of contract (as is the entire submission) and you as the manager are responsible for seeing that its pledges are honored.

STUDY COMMITTEES

Possible study committees include an executive board (consisting of you, your regulatory affairs liaison and the chief investigators), a safety board (consisting of at least two physicians and a biostatistician who are not directly involved in the conduct of the trials), a clinical events adjudication committee (consisting of at least three physicians who are specialists in the medical area and are not directly involved in the conduct of the trials), and other review panels specific to the investigation (e.g., pathology review, angiogram review).

APPENDIXES

In addition to a list of the journal and books that were cited in your proposal and a sample informed consent form, any or all of the following may be required depending on the requirements of your regulatory agency:

1. **List of investigators accompanied by copies of instructions provided to them**
2. **List of laboratories participating in the study accompanied by copies of instructions provided to them**
3. **List of personnel with access to treatment codes**

SAMPLE INFORMED CONSENT FORM

Introduction

Purpose of the Study

Randomization Procedure

- **If you agree to participate in this study, you will randomly (by chance) be assigned to either [describe treatment] or [describe alternative(s)]. You have a 50–50 chance to receive either treatment. Your physician will decide before randomization whether the addition of [name of proposed adjunct therapy] is advantageous.**

Procedure

Include only a brief description adopting the patient's point of view.

Potential Risks

- **The risks of this intervention [name] are very similar to [name the alternative intervention and describe the risks associated with it].**
- **(if warranted) As with any measurements of this type, there are certain risks which include dye or drug allergy, and. . . .**
- **(if applicable) The use of [name the adjunct therapy] is standard in all treatments of your condition and entail the following risks [list].**

Potential Benefits

- **Studies such as this are performed to determine the relative risk and benefits of these treatments. No definite benefits can be guaranteed by your participation in this study.**

Confidentiality

Assure the patients that their identity is protected.

Alternative Courses of Treatment

- **The following . . . and . . . are accepted standard treatments for your disease condition. If after consideration of these potential benefits and risks you do not wish to participate in the study, you**

and your doctor will decide which standard treatment may be appropriate for you.

Policy Regarding Research Related Injuries

- **In the event of injury resulting from your participation in this study, there will be no monetary compensation or subsidized medical treatment provided to you by any person involved in this research project including the study sponsor or (name of the institution).**

Payments or Additional Costs to Patients

- **There are no payments to patients participating in this study. The routine cost for this procedure will be billed to your insurance carrier. Nonroutine costs required by this study protocol (name them) will be paid by the sponsor.**

Problems or Questions

Patient's Consent

I have read the explanation about this study and have been given the opportunity to discuss it and to ask questions. I hereby consent to take part in this study.

Signature of Participant	**Date**
Signature of Investigator	**Date**
Signature of Witness	**Date**

PROCEDURES MANUALS

The design process is not complete until you have prepared a detailed list of the information that is to be gathered (see Chapter 9) and a manual of the procedures that are to be followed in gathering and evaluating the information.

We've already noted the variation that is inherent when we move from one patient to the next. With some interventions, there can be just as much variation in *how* the treatment is administered. With all interventions, there can be considerable variation in how the observations are made, even those on a single patient, unless we are consistent in the manner that we measure them. Moreover, if measurements are made on two different patients by two different investigators, then added to the normal patient-to-patient variation would be differences that result from differences in technique. The object of your procedure manual is to minimize these differences.

The manual should be the joint product of study physicians and professional writers. A separate manual should be prepared for each investigational laboratory and review panel.

Physician's Procedure Manual

After a brief reiteration of the purpose of the trials, the physician's manual should contain the following detailed instructions:

- **Determining eligibility. Define each criterion, explain its purpose, and note the requirement for strict adherence to. Give step-by-step instructions for obtaining the necessary information. For example, the eligibility questionnaire is to be completed by the patient initially and then is to be gone over item by item with the patient by the physician or the study nurse.**
- **Making and recording each of the necessary baseline and follow-up measurements. Note who is to take the observations (nurse, nurse practitioner, only the physician) and how measurements are to be recorded (include units). If materials are to be sent to an outside laboratory, describe how samples are to be collected and provide packing and shipping instructions.**
- **Administering the intervention. Preface by noting the following restrictions:**
  - **Experimental intervention can only be used with enrolled patients.**
  - **Intervention cannot begin until eligibility and specified baseline data have been submitted and an assignment made to treatment.**
  - **Intervention cannot be stopped or modified without first notifying you, the sponsor of the trials.**
  - **The only investigative procedures that may be used on an enrolled patient are those specified in the protocol.**

  Provide a step-by-step description of the procedure. If surgery is involved, the text should be accompanied by pictures and, preferably, by a videotape, compact disk, or DVD illustrating the procedure.
- **Reporting of adverse events. What events are likely; how they will be detected; how they will be summarized; how they will be reported.**

Include a section on patient compliance. Here is a possible wording. "Every patient contact should be used to educate, motivate, and reinforce compliance. Individualized caring attention by the members of your staff will also increase patient adherence to the prescribed therapy."

Incorporate any sample instructions you want the physician to pass on to the patient. For example, "One pill is to be taken twice a day with meals. In the event that you forget to take a pill at the designated time, please take it as soon as possible. If you miss a dose completely, please do not double up on the dose, but simply take your pill at the next regularly scheduled time. I'd also like you to make a record of any doses you do miss and to give me the list the next time you are in the office."

These instructions should be included as an appendix to the physician procedure manual in a form that lends itself to copying and distribution to patients.

Laboratory Guidelines

The laboratory guidelines prepared for each individual specialty laboratory should be as comprehensive as those provided to the physician. They should cover the preparation of the sample, shipping and receiving, and the particulars of the analysis and should be as detailed as possible.

This latter section should be written only after receiving a preliminary description from the laboratory. While it may seem foolish to create a document that merely parrots back the laboratory's own words, the result is to create a contract and, hopefully, to ensure uniformity in technique throughout the length of the trials.

INTERIM REPORTS

Your interim reports consist of all reports that are essential to successful conduct of the study. Included are reports on enrollment, submission of scheduled follow-up data, adverse events, and an ongoing abstract of findings.

TABLE 8.2 Interim Enrollment Report (1/6/03–1/7/03)

Site	Patients Enrolled	Enrollment Target	Prospect Interviews	Women Enrolled	Minority Enroll	Comments
002	8	17	11	2	0	
003	11	13	20	0	0	Must enroll women
004	0	6	0	0	0	Drop? Did not return call
005	5	20	10	0	5	TV ads begin Monday
006	10	15	15	2	3	
007	8	13	8	?	?	Some forms not entered
⋮	⋮	⋮	⋮	⋮	⋮	⋮
018	50	66	85	9	5	

Enrollment Report

> **As of 1 May 2003, 221 (52%) of the 400 anticipated patients are enrolled. This number includes 128 Caucasian, 26 African-American, 14 Hispanic, and 24 subjects designating themselves as Asian or other. A total of 856 potential subjects were interviewed. Sites 2, 4, and 8 have recruited 100%, 95%, and 80% of their expectations. Sites 5, 6, and 9 have not yet recruited any patients.**

In the hypothetical case depicted in Table 8.2, enrollment has passed the halfway mark, yet one of the sites has not recruited a single patient. What are you going to do about it?

Data in Hand

> **As of 1 August 2003, 338 initial follow-up reports from the 385 enrolled patients are in hand. Twenty reports are overdue including 5 from site 10. Laboratory results from 28 of these patients are missing or incomplete. (See Table 8.3.) One hundred and forty-five patients report partial or complete improvement on the self-diagnostic scale. Forty patients report worsening symptoms.**

TABLE 8.3 Angiogram Status and Follow-up Rate Tracking (3/3/04)

Site	Patients Enrolled	Follow-up Angio Not Done	Follow-up Missing	Patients for Analysis	Follow-up Rate*	Comments
002	17	0		17	100%	
003	13	5	2	6	46%	Core lab has all films
004	6	3		3	50%	
005	20	5		15	75%	
006	15	4		11	73%	
018	66	20	5	41	62%	Core lab has one; site is locating 3 others; one is unavailable.

Adverse Event Report

Three types of interim adverse event reports are of interest: (1) Frequency of specific adverse events, (2) number of adverse events per patient, and (3) adverse events by time period. If the trial is single-blind, all the above should be further broken down by treatment. The form of these reports is illustrated in Table 8.4 a, b, c. Distribution of the adverse events reports is highly restricted (see Chapters 13 and 14).

TABLE 8.4a Adverse Events by Treatment

	Current Date: 26 Sept. 2003	
Date of First Intervention 1 Aug 2003	With ABC	Without ABC
Q-wave MI	2	1
Non-Q-wave MI	5	10
Angina	1	1
Chest pains	8	4
Bleeding complications	4	1
⋮		
Other	3	5

TABLE 8.4b Adverse Events/Patient

Adverse Events	With ABC	Without ABC	Total
0	233	253	486
1	56	46	102
2	30	33	63
3	21	17	38
4	16	13	29
⋮	⋮	⋮	⋮
11	1	0	1
Total	300	294	594

TABLE 8.4c Adverse Events by Period

Period	With ABC	Without ABC	Total
Procedure	45	30	75
Two weeks	8	6	14
One month	39	24	63
⋮			
Six months	32	26	58
⋮	⋮	⋮	⋮
Total	300	294	594

Annotated Abstracts

The abstract is to be used for the initial report to the regulatory agency as well as a summary report for internal use. Breakdown by treatment is to be made only after the trials are concluded if the trials are double blind. The HTML format employed provides links to lists of missing data, protocol logic, and analysis software.

Prepared by Phillip Good, 30 July, 2001 using 12 July 2001 database
Company Confidential. For internal use only.

Among 705 patients, 28% of the 702 patients for whom data was available were females, and the mean age for the 702 patients for whom birth data was available was 63.9 ± 4.6.

- *Click here to see a list of patients with missing or erroneous dates.*
- *Click here to view program logic*
- *Click here to view SAS programs*

The 658 angiographic films analyzed showed average lesion length to be 12.5 ± 1.9 mm.

- *Click here to see a list of patients with missing film data.*
- *Click here to view program logic*
- *Click here to view SAS programs*

RVD was 2.76 ± 0.51 mm (n = 624).

- *Click here to see lists of patients with missing measurements.*
- *Click here to view program logic*
- *Click here to view SAS program*

Complex lesions were identified as long (≥12 mm but ≤32 mm) 54%, extra long (>32 mm) 4.3%, ostial 5%, and bifurcation 28% based on the results of 700 patients. Complex lesions were identified as CTO 4% based on the results of 699 patients.

- *Click here to see a list of patients whose eligibility forms are not on file.*
- *Click here to view program logic*
- *Click here to view SAS program*

Thirty-day MACE included 1 death, 3 Q-wave MIs, and 10 non-Q-wave MIs.

- *Click here to see a list of patients with missing film data.*
- *Click here to view program logic*
- *Click here to view SAS programs*

Postprocedural residual stenosis based on reports from 670 patients was 17.0% ± 2.1% in the traditional device group and 13.0% ± 1.4% in the novel device group.

- *Click here to see a list of patients with missing treatment assignments*
- *Click here to view program logic*
- *Click here to view SAS programs*

Binary residual stenosis based on an analysis of angiograms of the target lesion from 450 patients was 21.9% in the traditional group and 18.2% in the novel group.

- *Click here to see a list of patients with missing data*
- *Click here to view program logic*
- *Click here to view SAS programs*

Program Logic

Age
Age is determined for each patient by matching the procedure starting date with the birth date.

Lesion length
Average lesion length is calculated by taking the mean of the pre-procedural lesion lengths.

RVD
RVD is computed as the average of the actual not the interpolated preprocedural measurements.

Complex Lesions
The source of the information was the [prm] database.
Patients with no eligibility form on file were eliminated.
Records for events other than preprocedural or procedural were eliminated. If both records were present for the same patient, only the latter was used. Duplicate records were eliminated.

A lesion was

- long if $12 \le LGTH \le 32$; xtralong if $LGTH > 32$;
- ostial if segloc = 04
- bifurcation if lsnbch = 02 or bifurc = 04
- CTO IF event_id="PROC" AND pag_name="11_INITL" AND lsnocl=02 or 03.

MACE
Deaths derived from outcome database with otcm == 63.
Number of Q-wave MIs and non-Q-wave MIs derived from outcome database with cls == 9003 or 9004.
Time after procedure of adverse event was determined by subtracting the procedure starting date from the date of the event.

PRS
Postprocedure residual stenosis (PRS) is calculated from the postprocedure lesion MLD (the average of two measurements) and the postprocedure RVD.

BR
For those patients who had follow-up angiograms, binary restenosis (BR) is defined as restenosis ≥50% where restenosis is calculated from the follow-up target lesion MLD (the average of two measurements) and the RVD.

If angiography demonstrated binary restenosis at the target lesion between 14 days and 7 months and 2 weeks the patient was counted as having BR. Otherwise, a follow-up angiogram of the target lesion at eight ±0.5 months was used. If such an angiogram was not available, the first angiogram of the target lesion taken after four months was used as the basis of determination.

FINAL REPORT(S)

While it may seem curious that we would discuss the final report before we've even begun to collect data, the form of the final report(s) should be envisioned during the design process and only the numbers remain to be filled in during the analysis phase. Our reports should determine the nature of the data we collect. By preparing the form of the final report now, we have the opportunity to uncover any remaining discrepancies. It should be emphasized again that the quality of the final reports is heavily dependent on the clarity of the protocol.

The necessary reports include ongoing summaries (interim and final), a comprehensive report for the regulatory agency, and journal articles.

Regulatory Agency Submissions

The final report to the regulatory agency, like the interim abstract, should appear (and be) concise while providing links to detailed expositions. These latter should provide further links to summary tables and figures and, eventually, to the raw data itself. ICH (1996) should be consulted for overall guidelines.

Prefatory material should identify your project with its name, an alphanumerical ID if one was supplied by the regulatory agency, and your sponsor data so that the report can be immediately linked to the proposal that you submitted originally. Next should come a single summary paragraph outlining the form of the study, and stating the objective and the principal results. Here is an example:

> **400 patients with a prior history of . . . received a control treatment of . . . in a single-blind intent-to-treat study, while 385 patients**

> **received a . . . plus. . . . The binary restenosis rate of 23.2% for the control patients was significantly greater than the rate of 18.2% for the group receiving . . . plus. . . . The control group experienced more deaths and a statistically greater number of non-Q-wave MIs than those receiving the new treatment.**

Further paragraphs should be used to expound on the following topics:

- **Demographics. For example, "the experience of the males and females in the group were similar (see Table 1); those with fewer initial complications benefited the most from the new treatment (see Table 2)."**
- **Nature and frequency of adverse events, by type and treatment. Events should be classified by severity and by their relation to the intervention. Note whether a safety committee reviewed the events.**
- **Exceptions. Tabulate withdrawals, noncompliance, and modifications in treatment by starting treatment.**

Reproduce those sections of the protocol (updated to reflect any changes) dealing with end points, and your measurement and recording methods.

Tabular material, such as a table on the differing experiences of males and females, should be accompanied by a discussion of the statistical techniques employed and include measures of variability (23.2% ± 4.6%), of sample size ($n = 323$), and of statistical significance ($p < 2\%$). It also should include links to the data extraction and analyses programs so that the results may be independently validated.

Almost all the statistical analyses you perform will require you to first group the data. For example, you will want to combine the data from the various treatment sites, or from all patients regardless of sex or number of risk factors. You will have to precede your analyses by a justification of this grouping; that is, you will have to demonstrate there are no statistically significant differences in result among the categories you wish to group.[21] The templates you develop will form an essential part of your final report.

E-submission

> **A reviewer cannot refuse an electronic submission.—CDER Commitments, 1998**

---

[21] If significant differences exist among the categories, you will have to present separate tabulations for each distinct group. (For a more extensive discussion of this point, see Chapter 15.)

Your final submission to the regulatory agency will be in two parts: a printed copy of your final report and an electronic copy of your database and the programs you're used to retrieve and analyze the data. In the United States the latter is termed an e-submission or CANDA (Computer-Aided New Drug Application).

E-submissions, and computer-assisted product licensing applications submitted for vaccines and other biological products, shorten review time by reducing the need to sift through reams of paper to get the answers to reviewers' questions. Ordinarily, if a medical reviewer has a question about a specific patient in a drug trial, the reviewer sends a written request for the patient's record to the central document room where new drug application files are kept. It could take a day or more to get the information. Moreover, reviewing scientists often need to go back to you the drug sponsor for clarification or reworking of statistical data, which can delay the process for weeks. E-submissions eliminate much of this delay.

By employing computer-assisted data entry you are automatically in a position to submit electronically.

Between 1991 and 1994, e-submissions in the United States were approved about six months faster than traditional paper applications—in an average of 18.4 months compared with about 24.6 months. The FDA hopes that eventually all new drug applications will contain data that can be processed by computer. See also Agency Perspective on Electronic Submissions (as of 1999) at *http://www.fda.gov/cder/present/disom/index.htm*. For guidelines for electronic submissions in Europe, see *http://esubmission.eudra.org/*.

Whether or not, yours is an E-submission, you will still need to provide the regulatory agency with access to the database if requested and provide the agency with the details of the software used in the analysis.

Journal Articles

Drafting and publishing journal articles provides ammunition for your marketing department. They'll use quotes such as the following in your ads: "As reported in *JAMA*, or *Lancet*, or *Biotechnology Today*, our new product provides relief in 50% more cases." Your field representatives will give out copies of the articles to physicians as part of your product's information packet.

Four rules apply:

1. **Your article should mirror your report to the regulatory agency (though you may omit the material dealing with noncompliant patients and other exceptions).**

2. **Submit your article first to the top, high circulation journals and work your way down. (On the other hand, unrealistically high expectations and subsequent rejections may unnecessarily delay publication.)**
3. **Follow the guidelines for submission each journal provides.**
4. **"Describe statistical methods with enough detail to enable a knowledgeable reader with access to the data to verify the reported results."[22]**

Don't trust one of your investigators to prepare the article unless he or she has a track record of successful publication. Normally your marketing department should be able to provide a professional writer to work with the investigator. If not, hire one.

TO LEARN MORE

The Asilomar Working Group on Recommendations for Reporting Clinical Trials in the Biomedical Literature. (1996). Checklist of information for inclusion in reports of clinical trials. *Ann. Inter. Med*. 124:741–743.

Begg C; Cho M; Eastwood S; Horton R; Moher D; Olkin I, et al. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 276:637–639.

Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). *Guideline for Industry: Structure and Content of Clinical Reports*. 1996. *http://www.fda.gov/cder/guidance/iche3.pdf Fed. Reg.* (61 FR 37320). For guidelines on European submissions, see *http://www.eudra.ord/en_home.htm* and *http://medagencies.org*.

Good PI. (2000). *Permutation Tests*, 2nd ed. New York: Springer.

International Committee of Medical Journal Editors. (1991). Uniform requirements for manuscripts submitted to biomedical journals. *N. Engl. J. Med*. 324:424–428.

Kessler DA. (1997). Remarks by the Commissioner of Food and Drugs. *Food Drug Law J.* 52:1–5.

Long TA; Secic M. (1997). *How to Report Statistics in Medicine*. Philadelphia: American College of Physicians.

O'Connor M; Woodford FP. (1976). *Writing Scientific Papers in English.* Amsterdam: Elsevier.

Rutman O; Givens SV. (1997). Evolution of the CANDA at Roche. *J. Biopharm. Stat.* 7:605–615.

Switula D. (2000). Principles of good clinical practice (GCP) in clinical research. *Sci. Eng. Ethics* 6:71–77.

---

[22] Quote is that of the International Committee of Medical Journal Editors (1991).

Chapter 9

Recruiting and Retaining Physicians and Patients

RECRUITING PHYSICIANS

Your overall objective is to recruit physicians who can provide and care for the large number of eligible patients your study requires within the time period you've allotted for the study.

Ideally all trials would be conducted at a single site. This would keep costs to a minimum, ensure greater control over protocol administration, and eliminate the need for your statistician to correct for site-to-site differences. It's not going to happen. Still, you'll want to keep the number of sites to a minimum. Group practices, clinical research centers, and teaching hospitals are to be preferred to solo practitioners.

Obviously your first choice for the panel will be physicians you've worked with successfully on other projects. A measure of caution is needed even here. Sometimes such physicians turn into "contract professionals." They spend insufficient time with study patients or, because they are already participating in several other studies, may not have sufficient time to devote to yours. They may no longer be sufficiently stringent with regard to eligibility criteria. Or they may try to shape the results toward what they perceive as your expectations. Even with physicians you know well a preliminary on-site inspection is essential.

One can also try to recruit friends of "friends," that is, individuals referred by field representatives and existing investigators, but this procedure offers no particular advantages over a straightforward solicitation of all the investigators in a given area. Your focus should not be on friendship but on which physician's practice is likely to yield the most eligible patients.

A clinician on your staff should make initial contact with potential

RECRUIT WITH CARE

Thirteen leading medical journals will today warn that the promise of financial rewards is corrupting human clinical trials.

The editors will criticise pharmaceuticals companies for their use of private nonacademic research groups—called contract research organizations—CRO's—instead of scientists connected to universities and hospitals.
***Financial Times*, 10 September 2001**

A list of investigators who have repeatedly or deliberately failed to comply with FDA regulatory requirements for studies or have submitted false information to the study's sponsor is found at *http://www.fda.gov/ora/compliance_ref/bimo/dis_res_assur.htm*

investigators. McBride et al. (1996) report that mailing to individual physicians, a cumbersome and expensive method, has a very low response rate. Initial contacts with practice medical directors increase the participation rate substantially, and recruitment meetings with local practitioners improve both study participation and practice-project communication.

Other important factors in recruiting and retaining investigators according to Carey et al. (1996) are close liaison with local medical organizations, ongoing personal contact with the practices, and ongoing recognition of the value of the practicing physician's time.

Teaching Hospitals

Your second obvious choice, a teaching hospital staffed by research-oriented physicians who are almost guaranteed to see a large numbers of patients satisfying your inclusion criteria, has at least two major drawbacks:

The first is the patients themselves, a large proportion of who, being indigent, have had substandard or no medical care in the past, and thus present an entire constellation of symptoms and underlying etiologies in addition to those of primary interest.

The second is that academic physicians, professors, and fellows at medical schools and research institutions while understanding your need for a double-blind approach, thorough documentation, and patient consent forms, generally have experience only with smaller studies of more limited scope. Industry-sponsored trials often entail a 30–50 page protocol in contrast to the 5–10 page protocols common in the academic area, and industry collection forms can range from 50 to 200 pages per patient, in contrast to the 5–10 page forms of the typical academic trial. Academic research is typically completed in months rather than years, with the result that physicians drawn from teaching hospitals may prove less and less cooperative as trials continue, particularly if, in the later stages, it appears there are unlikely to be any publishable results.

Clinical Resource Centers

A third resource is profit-based clinical resource centers that have been set up specifically to conduct clinical trials. An annotated guide to centers in the United States may be found at *http://www.centerwatch.com/*.

A typical listing in this guide might note the number of trials the center has assisted in, the number of resident physicians, their backgrounds and clinical trials experience, and their medical specialties. One center in these listings advertises a single investigator supported by "five highly qualified clinical research coordinators, with an average of five years in pharmaceutical research each; three of the study coordinators are CCRC certified." As an additional bonus, this center provides "a database of over 12,465 clinical research volunteers. In addition, the center has access to a database of more than 35,000 patients from two private practices."

As you did when you hired a contractor to add a room to your home, probe well beyond the advertisement to establish the facts.

Look to Motivations

Let's suppose that you have gathered together a group of physicians who you feel are capable of recruiting and working with patients during the time period you have allotted for the study, that you have visited their offices and operating wards and are convinced they would make desirable members of the team. How can you persuade them to come aboard?

Consider the following eleven motivators listed by Spilker (1991):

1. **Enhance one's career (a priority for academic physicians, it can also function as a demotivator if the trial does not appear to be paying off)**
2. **Participate in scientifically exciting research**
3. **Obtain medical benefits for one's patients**
4. **Obtain new medical or scientific equipment provided by sponsors to enable trial (or purchased with monies available downstream)**
5. **Obtain new staff to help with clinical trial (this motivator could backfire as the physician begins to think of the staff as his own and wants to assign them to other tasks)**

6. **Obtain money that may be used for personal interests**
7. **Obtain money that may be used to conduct unsponsored trials of personal and professional interest**
8. **Publish scientifically and medically important journal articles**
9. **Develop a long-term relationship with you and your firm**
10. **Repay a favor (can only be pushed so far)**
11. **Be part of a team (this latter motivator is particularly important for physicians engaged in a solo practice or who, for a variety of reasons, feel estranged from their coworkers)**

Physician Retention

Don't go overboard on the sales pitch. No point in recruiting physicians who are not going to remain with the study or, worse, who remain on the payroll but do not contribute.

The first rule in successful retention is not to hire the wrong investigators to begin with. Some physicians may not be good candidates because of too strong a belief in one modality or another. Others may have been guilty of misconduct or of nonadherence to protocol in a prior trial.

No investigator should be brought on board without at least two interviews and a visit (not a telephone call) to the local medical society. Your sales representatives can be particularly helpful in providing feedback on a candidate's local reputation. You've read it once and now you get to read it again: *An on-site inspection is essential.*[23]

Follow up on your recruiting efforts to ensure retention of those you have recruited. Maintain ongoing personal contact with the practices. Constantly endeavor to show that you recognize the value of the practicing physician's time. And continue through newsletters, reports, and meetings to let the investigators know they are part of a team.

Get the Trials in Motion

On the one hand, you've been told repeatedly to get all your ducks in a row before you begin. On the other, you are most likely to lose physicians (and patients) who have signed up on paper but have not yet made an actual emotional commitment to the trials. Thus the majority of physician and patient recruitment should be performed only *after* the design process is completed and software development is well under way.

---

[23] Don't believe or don't want to believe that physicians can be corrupt? See Lock (1990) and Wells (1992).

Recruitment of physicians and the initial recruitment of patients should be brief and intense, concentrated in a few short weeks. Ideally training of the physician and his staff and appointment of a site coordinator should begin shortly thereafter. Physicians will begin to drop out or develop other interests if too much time elapses between recruitment and the start of the trials.

PATIENT RECRUITMENT

> **Successful recruitment depends on developing a careful plan with multiple strategies, maintaining flexibility, establishing interim goals, and preparing to devote the necessary effort.**[24]

Invariably the number of patients actually recruited is many times less than the number predicted. Physicians tend to overestimate the amount of eligible patients they will treat during the course of the study sometimes offering numbers 5 to 10 times in excess of what they might reasonably hope to achieve. Many investigators will recruit no patients at all, wasting the efforts you've put into training them and their staff, and pocketing any setup monies you've provided. Worse, some investigators will recruit exactly one patient (or one less than the number of treatment arms) with the result that their efforts are (almost) unusable. ("Almost" because any adverse events at that site would still need to be reported.)

To obtain the number of subjects you want in the time period you've allotted, you need to monitor the recruitment efforts of your investigators and engage in lengthy and costly recruiting efforts of your own.

Factors in Recruitment

Before going into the various methods for recruitment one might employ, let's first ask ourselves why a prospective subject might not enroll in our study. Four obvious reasons are that the subject

1. **Doesn't know about the study**
2. **Doesn't want to be in the study**
3. **Doesn't satisfy the eligibility criteria**
4. **Is enrolled in a similar study**

The first of these barriers can be addressed in three ways:

---

[24] Friedman, Furberg, and DeMets (1996, p. 141).

1. **Media campaigns—radio and TV announcements, newspaper and magazine articles directed at the population of prospective patients**
2. **Direct contact with all the physicians in a given area**
3. **Reinforcing the recruiting efforts of your study investigators**

All routes, not just the latter should be utilized.

Understanding the factors underlying the decision to participate doesn't take a rocket scientist. Put yourself in the potential subject's place: Why might you be willing to give up part of your time and place yourself at risk to be part of a clinical study? Humanitarian concerns, perhaps? We all want to be part of something meaningful.

But we are most likely to participate when we perceive the possibility of a direct benefit to ourselves or to those we love. For example, in a study of pain-relieving agents, individuals who have suffered from headaches or whose close relatives have suffered are more likely to be volunteers.[25]

The ideal advertising campaign will let prospective patients know they can satisfy their own needs while helping others. Other factors that tend to attract volunteers are the promise of personal attention and care by specialists (Mattson, Curb, and McArdle, 1985), money, and the lure of being part of something significant.

To increase participation in a study, you need to:

- **Understand what the benefits are**
- **Increase the benefits**
- **Ensure prospective study participants know about the benefits**

Importance of Planning

The importance of planning cannot be overestimated. You need to know whether other similar studies are in progress that might compete with yours for patients. A convenient registry of ongoing international (Australia, Britain, Canada, Hong Kong) clinical trials may be found at *http://www.controlled-trials.com/*. In the United States a similar registry may be found at *http://www.centerwatch.com/patient/trials.htm*. I'd also recommend you conduct trials of the various recruitment methods to see which is likely to be the most successful.

Tilley and Shorack (1990) found typical problems that arise during recruitment include the following:

---

[25] But see Ellis et al. (1999).

- **Inadequate funding for screening**
- **Unwillingness of physicians to refer patients**
- **Overestimation of the prevalence of the condition**
- **Overly rigorous entry criteria**

All of these barriers to success can be addressed to a degree by careful planning. The good news is that computer-assisted data entry gives you the opportunity to continuously monitor recruitment and respond quickly.[26]

Ethical Considerations

Enrollment must be monitored on a continuous basis to forestall the tendency of the very few less than ethical physicians to enroll unsuitable subjects or to skip (or, more often, sidestep) informed consent. Ideally eligibility forms should be reviewed on the day they are completed and prior to the start of the trial itself. (The notable exception being when immediate intervention is dictated.) Fortunately, computer-assisted data entry facilitates rapid review. But computerized analysis alone is not adequate. Frequent visits should be made routinely to each investigator's site.

A word of caution. While investigator's often rely successfully on referrals from potential subjects' personal physicians, Sugarman et al. (1999) warn that direct solicitation of subjects by their personal physician does not increase the recruitment rate and may be unethical.

Mass Recruiting

Mass recruiting efforts may be directed both to the community at large and to physician-practice based populations. One way to reach the community at large is by providing free screening at health fairs, church groups and other organizations, and sports events. Other alternatives in common use include mass mailings with utility bills, and newspaper and radio advertisements. No one best method exists. Given the innumerable specialized contacts required with any of these methods, I'd recommend employing a professional recruiting firm.

Handberg-Thurmond et al. (1998) found the highest yield was obtained by screening records of patients directly referred by a physician for possible study entry. Working through physicians, Margitic et al. (1999) found the self-administered office-based questionnaire to be the least costly strategy for one site ($14 per randomized participant), followed by patient mailing at another site ($58). The direct telephone contact method utilized at one site serving primarily a minority population yielded a cost per randomized participant of $80.

---

[26] You'll find a list of commercially available software for use in monitoring in the Appendix. Forecasting methods are described in Chapter 14.

PATIENT RETENTION

All your recruiting efforts will go for naught (and double your costs) if the patients you've recruited drop out of the study or do not comply with the protocol.

The three keys to patient retention are:

- **Selecting the appropriate participants**
- **Optimizing the trial experience from the patient's point of view**
- **Monitoring compliance**

The more reasons a patient has for participating in a trial, the more likely he is to remain in compliance. Money alone won't cut it, at least not over an extended period. Though there has been extensive research in the area, no single set of demographic factors has proved to be consistent predictors of success. (See, for example, some of the articles in Haynes, Taylor, and Sackett, 1979, and in Schumaker, Schron, and Ockene, 1990.)

Keep the interval between screening and the actual onset of treatment to a minimum. Until the actual onset of treatment, prospective patients feel little or no loyalty to the study. Consciously or unconsciously they may have already forgotten (or may actually regret) their decision to enroll. When the delay is protracted, the prospective patient should be contacted and, if possible, rewarded during the interval to reinforce their commitment.

Ensure the patient is fully informed (Sturdee, 2000) and that the directions they are provided with are unambiguous. Study physicians should be encouraged to spend ample time with each patient for this purpose. Your team should prepare and provide the physician with descriptive materials to be given the patient at the conclusion of an interview.

Physicians should be encouraged (and paid) to give all study patients VIP treatment. No long delays in crowded waiting rooms or being left for hours half-dressed in some isolated chamber. If a wait is necessary—surgeons do have emergencies—the site coordinator may have to spend time with the patient or make immediate alternate arrangements. In any event, the health provider, the site coordinator, and all members of the provider's staff who come in contact with the patient must be prepared to spend additional time to provide for the trial subject's need for assurance.

Not all of a patient's declared needs are genuine; some, from the health provider's point of view, are spurious. A major motivator for having a sponsor-paid coordinator at each site is to have someone to deal in a positive fashion with the overly garrulous, the overly demanding. We not only want the patient, but the investigator and his staff, to remain committed throughout the study.

Ongoing Efforts

None of the preceding can be relied upon to ensure compliance unless you continuously monitor the investigators and prod them, if necessary, to submit scheduled follow-up reports. Again, computer-assisted data entry can only facilitate such monitoring.

Your staff can contribute to compliance in several additional ways:

- **Keep the protocol simple. One pill a day is preferable to three- or four-times-a-day regimens. Crossover designs will only confuse the participants if they entail changes in dosage schedules. Keep it simple.**
- **Provide special pill dispensers or automate injectibles so that the patient has less to keep track of. Colorful reminder stickers, a watch that says "it's time to take your pill," and rewards for compliance are also of value.**
- **Consider preparing and distributing a newsletter to and for study subjects that will make them feel part of a team engaged in a worthwhile effort. The newsletter should be distributed on at least a quarterly basis (and more often if there is actual news). The possible contents of such a newsletter might include current statistics on the study's progress, profiles of study investigators, and reports gleaned from the media on the disease condition that is at the core of the study.**

Finally, you need to develop a method for monitoring patient compliance. Automated methods and pill counts are of dubious value. (See the texts cited earlier.) I recommend direct contact, via telephone, with each participant. Not that the information you gather will be any more reliable, but such calls serve the dual purpose of making the patient feel part of something important and thus more conscientious in compliance.

You may learn via telephone of adverse events: "I had to stop taking the pills because I was throwing up." Such responses should trigger calls from your staff to the subject's physician for further investigation. As the "blindness" of the treatment must be preserved, those making the calls on your behalf should not be aware of which treatment arm the patient was assigned to.

Run-in Period

Opinions differ on whether a run-in period should be used to identify and exclude patients who are unlikely to remain in compliance. During this period, best limited to three to six weeks prior to the start of the actual intervention, potential participants would be given either active medication or placebo and their compliance monitored. (See sidebar)

According to Friedman, Furburg, and DeMets (1996), the results have been almost uniformly positive—see, for example, Knipschild, Leffers, and Feinstein (1991) and Lang (1990). Although single-blind placebo run-ins are in common use today, Evans (2000) finds the practice ethically unjustified if they entail the withholding of medication. Milgrom et al. (1997) argue that run-ins have not yet been shown to be cost effective and may endanger recruitment success.

Davis et al. (1995) put the hypothesis to the test by prescribing placebo for a three-week period prior to the start of their clinical trials, but not using the results of this run-in period as an entry criterion. Of the 431 participants in their study, 66 (15%) who took less than 80% of the prescribed placebo or who failed to return their unused placebo pills were classified as poor run-in adherers. Poor run-in adherence was associated with lower educational attainment. At 3 and 6 months of follow-up, mean adherence was 89.3% and 83.4% among all participants. Exclusion of poor run-in adherers would have increased these means by one to two percentage points to 90.9% and 85.5%, respectively. Would a one to two percentage point gain in enrollment be worth the expense of those three extra weeks?

RUN-IN PERIOD

The following material is taken from the protocol for a study of colorectal adenoma chemoprevention using aspirin:

At six weeks, assess by telephone the patient's suitability for randomization based on compliance, motivation, and toxicity. If the patient appears suitable, telephone them again in 10 weeks.

If the patient appears suitable at 10 weeks, proceed to randomization. If the patient is deemed not suitable at 6 or at 10 weeks, complete and submit the eligibility form stating the reason for withdrawal.

Run-in periods also can be used to exclude placebo responders and subjects who cannot tolerate or do not respond to active drug. Pablos-Mendez, Barr, and Shea (1998) argue that the result of such use is to select a group of individuals who may differ markedly from patients undergoing active clinical management for this problem. But after reviewing some 101 studies, Trivedi and Rush (1994) find the use of a run-in period does not appear to enhance the drug/placebo differential.

Schechtman and Gordon (1993) find that run-in strategies are most likely to be cost effective under the following conditions: (1) per patient costs during the postrandomization as compared to the screening period are high, (2) poor compliance is associated with a substantial reduction in response to treatment, (3) the number of screened patients needed to identify a single eligible patient is small, (4) the run-in is inexpensive, (5) for most patients, the run-in compliance status is maintained following randomization, and (6) many subjects excluded by the run-in are treatment intolerant or noncompliant to the extent that little or no treatment response is expected. Schechtman and Gordon find that run-ins are least cost effective if their only purpose is to exclude ordinary partial compliers.

BUDGETS AND EXPENDITURES

Sweat the small stuff. Attornies do it; so do CPAs. Don't just record major expenditures—air travel and outlays to advertising agencies—but track every phone call and the time your staff spends on it. Successful recruiting, and success is defined as retaining those you recruit, requires sustained effort. See Chapter 15.

TO LEARN MORE

Agras WS; Bradford RH; Marshall GD, eds. (1992). Recruitment for clinical trials: The Lipid Research Clinics Coronary Primary Prevention Trial experience. Its implications for future trials. *Circul.* 66 Suppl. IV: IV-1–IV-78.

Carey TS; Kinsinger L; Keyserling T; Harris R. (1996). Research in the community: recruiting and retaining practices. *J. Comm. Health*. 21:315–327.

Cramer JA. (1998). *Improving and Supporting Patient Recruitment In Clinical Trials Conducted by Academic Medical Centers*. Oak Brook, IL: University Health System Consortium.

Davis CE; Applegate WB; Gordon DJ; Curtis RC; McCormick M. (1995). An empirical evaluation of the placebo run-in. *Control. Clin. Trials* 16:41–50.

Ellis PM; Dowsett SM; Butow PN; Tattersall MH. (1999). Attitudes to randomized clinical trials amongst out-patients attending a medical oncology clinic. *Health Expect*. 2:33–43.

Evans M. (2000). Justified deception? The single blind placebo in drug research. *J. Med. Ethics* 26:188–193; comment 26:477.

Gillum RF; Barsky AJ. (1974). Diagnosis and management of patient noncompliance. *JAMA* 228:1563–1567.

Haidich AB; Ioannidis JP. (2001). Determinants of patient recruitment in a multicenter clinical trials group: trends, seasonality and the effect of large studies. *BMC Med. Res. Methodol.* 1:4.

Handberg-Thurmond E; Baker A; Coglianese ME, et al. (1998). Identifying high yield sources of patients with coronary artery disease for clinical trials: Lessons from the Asymptomatic Cardiac Ischemia Pilot (ACIP) experience. The ACIP Study Group. *Clin. Cardiol*. 21:176–182.

Haynes RB; Taylor DW; Sackett DL, eds. (1979). *Compliance in Health Care.* Baltimore: Johns Hopkins Press.

Hunninghake DB; Blaskowski TP, eds. (1987). Proceedings of the workshop on the recruitment experience in NHLBI-sponsored clinical trials. *Control. Clin. Trials* 8 Suppl. 6S–30S.

Keith SJ. (2001). Evaluating characteristics of patient selection and dropout rates. *J. Clin. Psychiat*. 62 Suppl 9:11–14; discussion 15–16.

Knipschild P; Lffers P; Feinstein AR. (1991). The qualification period. *J. Clin. Epidemiol.* 44:461–464.

Lang JM. (1990). The use of a run-in to enhance compliance. *Stat. Med.* 9:87–95.

Lock SP. (1990). Research Fraud: Discouraging the others. *BMJ* 301:1348.

Lung Health Study Research Group. (1993). The Lung Health Study: design of the trial and recruitment of participants. *Control. Clin. Trials* 14 Suppl.

Margitic S, et al. (1999). Challenges faced in recruiting patients from primary care practices into a physical activity intervention trial. *Prev. Med*. 29:277–286.

Mattson ME; Curb JD; McArdle R; the AMIS and BHAT Research Groups. (1985). Participation in a clinical trial: the patients' point of view. *Control. Clin. Trials* 6: 156–157.

McBride PE; Massoth KM, et al. (1996). Recruitment of private practices for primary care research: Experience in a preventive services clinical trial. *J. Fam. Pract*. 43:389–395.

Milgrom PM; Hujoel PP; Weinstein P; Holborow DW. (1997). Subject recruitment, retention, and compliance in clinical trials in periodontics. *Ann Periodontol*. 2:64–74.

Pablos-Mendez A; Barr RG; Shea S. (1998). Run-in periods in randomized trials: Implications for the application of results in clinical practice. *JAMA* 279:222–225; comments 279:1526–1527.

Ruffin MT 4th, Baron J. (2000). Recruiting subjects in cancer prevention and control studies. *J. Cell Biochem.* 34 Suppl:80–83.

Schechtman KB; Gordon ME. (1993). A comprehensive algorithm for determining whether a run-in strategy will be a cost-effective design modification in a randomized clinical trial. *Stat. Med*. 12:111–128.

Schumaker SA; Schron EB; Ockene JK, eds. (1990). *Health Behavior Change.* New York: Springer.

Spilker B; Cramer JA. (1992). *Patient Recruitment in Clinical Trials*. New York: Raven Press.

Sturdee DW. (2000). The importance of patient education in improving compliance. *Climacteric*. 3 Suppl 2:9–13.

Sugarman J; Regan K; Parker B; Bluman LG; Schildkraut J. (1999). Ethical ramifications of alternative means of recruiting research participants from cancer registries. *Cancer* 86:647–651.

Tilley BC; Shorack MA. (1990). Designing clinical trials of treatment for osteoporosis: Recruitment and followup. *Calcif. Tissue Int*. 47:327–331.

Trivedi MH; Rush H. (1994). Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacol*. 11:33–43; comments 15:105, 107.

Wells F. (1992). Fraud and misconduct in clinical research: Is it prejudicial to patient safety? *Adverse Drug React. Toxicol. Rev*. 11:241–255.

Chapter 10

Computer-Assisted Data Entry

Electronic Case Report Form (e-CRF) means an auditable electronic record designed to record information required by the clinical trial protocol to be reported to the sponsor on each trial subject.

Guidance for Industry, Computerized Systems Used In Clinical Trials[27]

Computer-assisted data entry offers at least six advantages over paper case report forms:

- **Immediate detection and correction of errors. Mistakes such as typographical errors and misplaced decimal points are detected and corrected at the time of entry. No data is lost as a result of lapses in memory.**
- **Reduced sources of error. Eliminating the need to recode and reenter case report forms eliminate two potential sources of error. There is a corresponding reduction in costs.**
- **Open-ended reporting. If inspection of the "other" category reveals that "protein imbalance" is being written in with a relatively high frequency, then "protein imbalance" can be added to the options on the pull-down menu. Printed case report forms are fixed, lifeless.**
- **Quality control. Even the best-designed forms can contain ambiguities just as even the best-designed trials can have unexpected consequences. Measuring devices can go out of tolerance. By tabulating and monitoring the information as it is collected, problems at some or all of the sites can be detected and corrected early on.**
- **Improved investigator relations. If a trend is detected, particularly if it only involves one or two sites, it can be difficult to communi-**

---

[27] See *http://www.fda.gov/ora/compliance_ref/bimo/ffinalcct.pdf.*

cate the need for modification of procedures without offending some investigators. Let the clinical research monitor (CRM) and the investigator jointly blame the software.
- **Ease of access. Generally, the same software that simplifies data entry makes it easy for the noncomputer professional to access and display the result. (We expand on this point in the next chapter.) Both your staff and the regulatory agency will have earlier access to trial data compared with paper CRFs.**

Many regulatory agencies such as the FDA now accept and even prefer electronic submissions, thus doing away with the need to manage or store paper case report forms. If paper forms are required, they are readily produced. And if a paper form turns up missing, it is easily regenerated from the electronic record and submitted to the investigator for signature. (Security procedures for electronic records are discussed in the next chapter, also.)

Implementation of computer-assisted entry involves three steps:

1. **Developing and testing the data entry software**
2. **Training medical and paramedical personnel in the software's use**
3. **Monitoring the quality of the data**

We discuss the first two of these steps in the following sections, and the last step in Chapters 13 and 14.

PRE-DATA SCREEN DEVELOPMENT CHECKLIST

All required data are grouped by the individual who collects the data (patient, front-office person, nurse, physician) and according to the time at which it will be collected (initial screen, baseline, one-week follow-up).

For each data item, the units and acceptable range are specified; see Table 10.1.

Develop the Data Entry Software

The first steps in software development are as follows:

- **Decide which software product to use to develop the data-entry screens. (A list of commercially available software is provided in the Appendix.)**
- **Organize the required information into functional groups.**
- **Consult the CDISC Submission Standards.**
- **Prepare a flowchart or Gant chart for the development process.**

The responsibilities for choosing the development languages for data entry, data management, and data analysis are normally divided

TABLE 10.1 Data Specifications Table

Item	Group	Units	Question If	Reject Unless
Year of birth	Bp	Year		17 < (Current year – Birth year) < 81
Diastolic pressure	B, Fn	mm/HG	DP < 50 or DP > 110	30 < DP < systolic pressure

among the lead developer, the data manager, and the statistician. The project manager may be called upon to resolve conflicts not only among the members of this committee but with other units of the corporation (see sidebar). We've provided below a set of guidelines for software selection and a list of commercially available choices in the Appendix.

The lists of required information and the associated questions prepared by the design committee should be divided into functional groups. Each group consists of a set of questions that will be answered at the same time by the same individual. These groupings should parallel the time line you develop during the design phase.

- **Eligibility**
    - **Questions to determine eligibility for inclusion in the study**
    - **Patient demographics including risk factors**
- **Baseline**
    - **Evaluation of condition**
    - **Laboratory values**
    - **Special studies (e.g., angiogram)**
    - **Concurrent medications**
- **Intervention data**
- **(If applicable) Hospital summary**
- **Follow-up**
    - **Evaluation of condition**
    - **Events during interval**
    - **Laboratory values**
    - **Special studies (e.g., angiogram)**
    - **Concurrent medications**
- **Adverse event reports**
    - **Nature of event**
    - **(When applicable) Hospital summary, special studies, autopsy**
- **Protocol deviation**

Obviously each type of special study will require its own set of data entry screens. Normally one of the CRMs will oversee preparation of these groupings.

The lead developer is responsible for preparing a flow- or Gant chart for the development process. This chart will include the work assignments for each individual assigned to the project. I recommend that each functional group be the responsibility of a single developer working in tandem with a single CRM. Between them they will work out the context and sequencing of the screens needed to record their portion of the data.

One natural ordering of tasks follows the sequence in which the screens will be completed at the study centers. Those screens devoted to eligibility determination that contain the inclusion and exclusion criteria should be developed first, followed by the screens that will contain the baseline clinical information, risk factors, medical history, physical assessment, current medications, and baseline laboratory values. For the reasons outlined in Chapter 7, these screens should already be tested and in the hands of the investigators while the last of the follow-up, adverse event, and patient contact forms is still undergoing development.

Avoid Pre-defined Groupings

Avoid the use of pre-defined groups in forms.

For example, rather than ask the patient to classify their smoking habit as in Smoker (never, quit over 1 month, $<\frac{1}{2}$ pk/day, $\frac{1}{2}$ to 1 pk/day, >1 pk/day); have them enter the number of years they've smoked, their average pack per day consumption, and whether they are current smokers.

Rather than classify cholesterol levels as in Hypercholesterolemia (<200 mg/dl, 200 to 235 mg/dl, requires medication), enter the exact measurement of cholesterol level obtained in baseline screening.

Avoiding pre-defined groupings gives us much greater flexibility and allows us to use metric variables rather than categorical ones, paving the way for the use of more sensitive statistics. We can measure exposure to cigarette smoke in pack years or we can classify smokers as never, quit over 2 months, $<\frac{3}{4}$pk/day, $\frac{3}{4}$ to 2 pk/day, >2 pk/day.

CDISC SUBMISSION STANDARDS

The standards have five primary goals:

1. **Simplify and speed up the regulatory process by communicating clear descriptions of the structure, purpose, attributes, and contents of each dataset and dataset variable.**
2. **Simplify and speed up the regulatory process by providing the regulator with data structures with which he is already familiar.**

SAMPLE FORM SPECIFICATIONS

Form: Risk Factors 1
To be completed at: Baseline Patient Interview
To be completed by: Examining Nurse

FIELDS
Patient Name (last, first, MI)
Patient ID (display only)
Patient address and telephone number (display/update)

Does patient have significant GI bleeding (yes/no)?
Does patient have peripheral vascular disease (yes/no)?

Diabetes mellitus (none, treated with exercise diet alone, oral hypoglycemics, insulin)
Current smoker (yes/no)
Smoker (current or past) ______number of years; ______number packs per day
Hypertension (<90 mg Hg, 90–100 mg Hg, requires medication)

Has patient had a previous myocardial infarction? (yes/no)
(skip next fields if no) Date of most recent MI
Q-wave (yes/no/unknown)

Weight (specify Kg or lb)(question if not 100 to 280)(refuse if not 80 to 325)

Specifications prepared by: L Moore 19 Nov 2002
Specifications approved by: JR Moon 8 Dec 2002

3. **Allow reviewers to replicate most analyses, tables, graphs, and listings with minimal or no transformations.**
4. **Enable reviewers to easily view and subset the data used to generate any analysis, table, graph, or listing without complex programming.**
5. **Facilitate the combination of data from multiple sources. For example, if your company were to acquire or be acquired by a second firm employing CDISC, the clinical trial data collected by the two firms could be readily merged.**

No need for you and your staff to reinvent the wheel. Approximately 80% of the data you collect is covered by the CDISC guidelines. The guidelines also identify minimum core data elements that normally should be included with basic safety datasets. Also provided is some indication of the frequency with which data should be collected.

Here is an example:

> **Format of variables for similar types of data should also be consistent within and across studies. For example, all calendar dates (e.g., date of birth, screening visit date, randomization date, date of death) should use the same format for representing the date. The standard recommended is ISO 8601, which has two formats for dates represented as year, month, and day. The extended format of YYYY–MM–DD (SAS format: yymmdd10) is recommended, although the basic format of YYYYMMDD (SAS format: yymmddn8) may be used if space is more critical than human readability. Likewise for time, HH:MM is preferred over HHMM. More information about ISO (International Organization of Standards) is available at *http://www.iso.ch/*.**

For details, download the CDISC Submission Metadata Model and Introduction to the CDISC Submissions Data Domain Models from *http://www.cdisc.org/publications/index.html.*

SCREEN DEVELOPMENT

In computer-aided data entry, the computer's screen, approximately 80 characters wide by 24 lines, plays the role that printed case report forms once did. There is no need to copy or ape the printed form. The focus should be on making effective use of the screen. For example, rather than trying to cram a single form onto a single screen, the layout should be dictated by the comfort and convenience of the potential user.

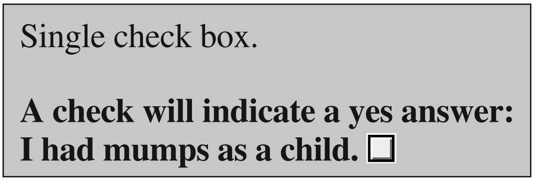
While the developer is responsible for the layout, the CRM should dictate the sequencing of questions and screens based on his or her knowledge of how the potential user (nurse, technician, specialist) is likely to acquire the information. The CRM is also responsible for filling in any gaps left by the design committee when they specified the range of permissible answers for each question.

An example would be a question on smoking habits. To the selection, "a pack a day," "two packs a day," "more than two packs," the CRM might need to add, "less than a pack a day."

Each question is represented on the screen by one of three formats, the radio button and pull-down menu for multiple-choice questions and the type-and-verify field for numeric responses.

Radio Button

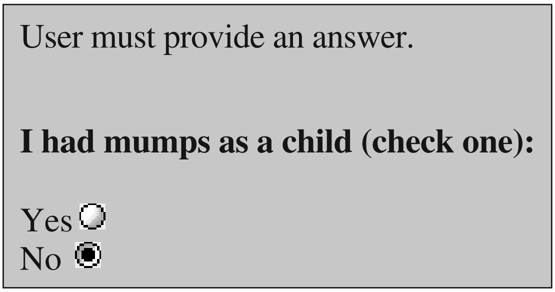
The radio button depicted in Figure 10.1*a* is recommended when there are only a few options and only one option may be selected. All alternatives should be displayed. A single check "yes" button as



Single check box.

A check will indicate a yes answer:
I had mumps as a child. ☐

FIGURE 10.1a

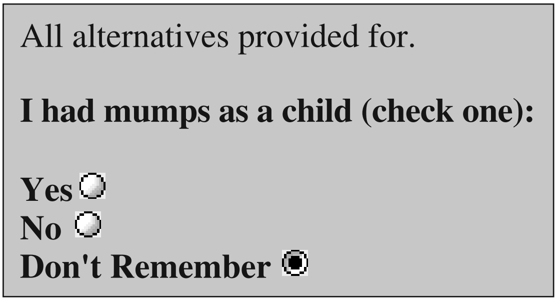


User must provide an answer.

I had mumps as a child (check one):

Yes ○
No ●

FIGURE 10.1b

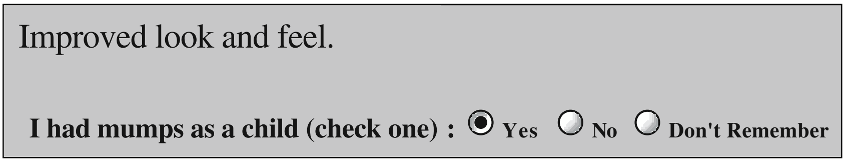


All alternatives provided for.

I had mumps as a child (check one):

Yes ○
No ○
Don't Remember ●

FIGURE 10.1c



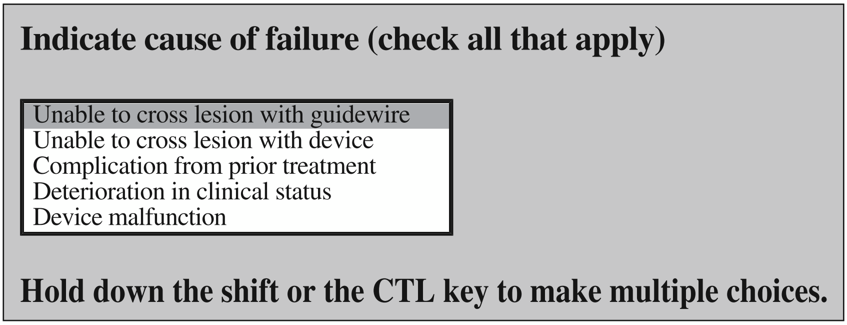
Improved look and feel.

I had mumps as a child (check one) : ● **Yes** ○ **No** ○ **Don't Remember**

FIGURE 10.1d

in Figure 10.1a is not acceptable. Figure 10.1*b* shows the correct approach. If neither "yes" nor "no" is checked, the cursor will not advance to the next question.

What if the respondent doesn't know or doesn't remember the answer? Then a third option should be incorporated as in Figure



Indicate cause of failure (check all that apply)

Unable to cross lesion with guidewire
Unable to cross lesion with device
Complication from prior treatment
Deterioration in clinical status
Device malfunction

Hold down the shift or the CTL key to make multiple choices.

FIGURE 10.2

10.1*c*. Skipping the question cannot be permitted, for a major objective of computer-assisted data entry is the elimination of missing data and the need for extensive time-consuming follow-up.

Figure 10.1*d* illustrates the use of graphics and layout options to create a user-friendly design for the data entry screen.

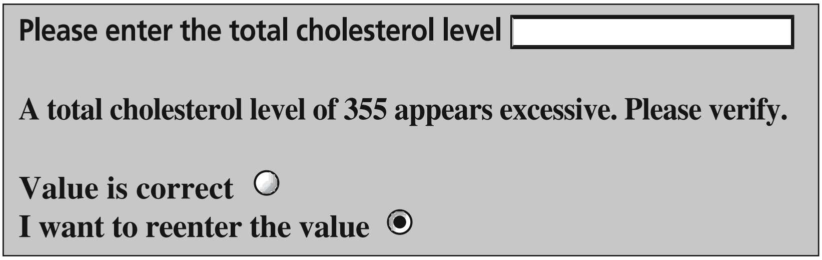
Pull-down Menus

Pull-down or pop-up menus are of two types, those that permit only a single selection from a menu of choices and those that permit multiple selections. The type of permission needs to be specified in advance by the forms-design committee.

Note in Figure 10.2 that not all the choices are displayed but can be accessed by scrolling through the pull-down menu using the side arrows. A field labeled "other" would be in the part of the menu we can't see.

Type and Verify

The type and verify field (Figure 10.3) is used for two types of data. For measurements and for comments such as "other risk factors include. . . ." A set of bounds needs to be specified for each measurement that will be entered in a type-and-verify field. Actually two sets of bounds need to be specified; the first set rules out the impossible, a negative value of cholesterol, for example. If an impossible value is displayed, the following message would appear on the screen, "A negative value is not possible, please reenter the value. Press enter to continue." When the user presses the enter key, the cursor returns to the field where the erroneous entry was made so that data can be reentered. The second set of bounds delineates so-called normal values, a total cholesterol level of more than 100 or less than 250, for example. Checking a "yes" would confirm the entry; checking a "no" would return the cursor to the field where the erroneous entry was made.



Please enter the total cholesterol level

A total cholesterol level of 355 appears excessive. Please verify.

Value is correct ○
I want to reenter the value ●

FIGURE 10.3

GUARANTEEING FAILURE

A sure way to guarantee failure is with bizarre keypunch instructions. Bumbling's printed case report form listed nine possible adverse events (including an "other" category). Thus question 17.4 was Myocardial infarction, yes or no; question 17.5 was Stroke, yes or no, and so forth. The secret to analyzing the data was to realize that all nine questions had been encoded to a single field using a total of 12 codes, listed—by the time I caught up with the ill-fated project—only on a faded handwritten piece of paper.

To discourage casual users from attempting to scan the database by eye, Bumbling made sure a different set of codes would be used on each new form. While an atherectomy might be coded as a 420 on the adverse event form under the heading "action taken," when the atherectomy was actually performed it would be coded on the repeat revascularization form as a 511.

Confused? So was everyone connected with the project.

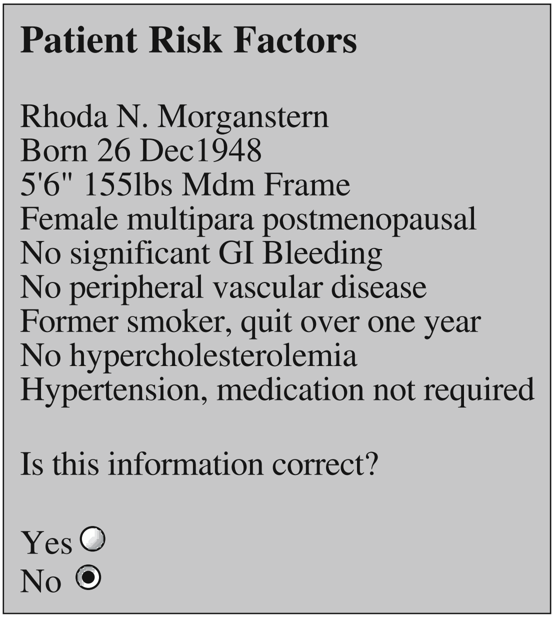
When the Entries Are Completed

After each screen is processed, a summary of the entries is displayed as in Figure 10.4, along with the message, "Are these entries correct, Yes or No?" A "yes" answer results in storing the entries in a file on disk and advancing the display to the next screen. A "no" answer returns the display to the just-completed screen so that corrections can be made.

Completing and accepting the last screen in a functional group triggers a printout of the completed case report form.

Audit Trail

One ought to have as much or more confidence in the data derived from computerized systems as if the data had been originally in



Patient Risk Factors

Rhoda N. Morganstern
Born 26 Dec1948
5'6" 155lbs Mdm Frame
Female multipara postmenopausal
No significant GI Bleeding
No peripheral vascular disease
Former smoker, quit over one year
No hypercholesterolemia
Hypertension, medication not required

Is this information correct?

Yes ○
No ●

FIGURE 10.4

paper form. Some guiding principles for maintaining data integrity and a clear audit trail where computerized systems are being used to create, modify, maintain, archive, retrieve, or transmit clinical data may be downloaded from *http://www.fda.gov/ora/compliance_ref/bimo/ffinalcct.htm.*

ELECTRONIC DATA CAPTURE

Electronic Case Report Forms (e-CRF) are just one facet of electronic data capture (EDC). The others include:

- **Direct data acquisition from laboratory instruments**
- **Handheld devices that allow patients and their caretakers to enter symptom/treatment data electronically accompanied by an automatic time-date stamp.**

The only essential information that continues to elude EDC is *interpretation*. For example, “Tissue is malignant,” “EKG reveals a myocardial infarction,” “Spot on the mammogram is a cyst,” “Adverse event is treatment related.” Interpretations must first be recorded on paper along with a dated signature before being entered into a clinical database.

CODING FOR CHAOS

Bumbling Pharmaceutical's Information Services Director had joined the company in an era when expanding memory was done in chunks of kilobytes rather than megabytes and a large hard disk was one that held 10 megabytes instead of 5. Determined to save computer memory, he ruled that information should be coded whenever possible.

The original printed case report form had provided for separate entries of each of a half dozen risk factors, with each factor further broken down into subcategories. Smoking history, for example, was broken down into "never smoked," "former smoker," and "current smoker." In the course of recoding the data, each category was assigned a separate numeric value so that "never smoked" was coded as 000, and "former smoker" as 021. All the "no's" on the form were assigned the same value of 000. The results were disastrous.

The designers of the form has assumed a 000 would appear on the completed form only if the patient answered "no" to all questions. But they had neglected the possibility of missing data. If the examining physician omitted to record whether or not the patient had diabetes, and checked "no" to all the other questions, a 000 appeared in the database implying the patient did not have diabetes, even though quite the opposite might be true.

TESTING

Testing is the responsibility of every team member and not just the testing department. Quality should be built in from the start. If multiple developers are employed, frequent meetings are necessary to ensure that the developers use common naming and programming conventions. Each screen should be tested separately by its developer, then tested again by the developer as part of the larger integrated package before the program is turned over to the testing group to repeat the entire process.

The purpose of testing is twofold. First is to ensure that the program does what it is supposed to do. If an 11.6 is entered from the keyboard, 11.6 should be recorded in the file and not 11.8 or 116. A cholesterol level of 250 should trigger a warning message as shown in Figure 10.3. A user should not be able to advance to the next screen of a series without filling in answers to all the questions on the screen she is currently viewing.

The second purpose of testing is to ensure that the program does *not* do what it is *not* supposed to do. If a cholesterol level of 2500 or 2.5 is typed in, it should not be entered into the file. And, most important, a doctor or nurse should never find herself staring at a screen that displays the following:

```
DataEntry32 caused an invalid page fault in
module MFC42.DLL at 015f:5f4040fd.
Registers:
EAX = 00000000 CS = 015f EIP = 5f4040fd EFLGS = 00010246
EBX = 00000000 SS = 0167 ESP = 007cf880 EBP = 008f2870
ECX = 5f4d1b4c DS = 0167 ESI = 007cf8a0 FS = 1147
EDX = 00000006 ES = 0167 EDI = 008f2504 GS = 0000
Bytes at CS:EIP:
83 78 f4 00 0f 8c 7c 7b 05 00 8b ce e8 6b f8 ff
Stack dump:
00000000 008f33a8 0045f8ca 008f2504 007cfb40
007ec8dc 008f3130 008f3130
5f4d1b58 008f2870 00000005 008f3130 008f3130
007cf8a8 007cf944 007cf944
```

No, I don't know what these numbers mean—does anyone?—but I know that once I see a display like this, I can wave goodbye to all the work I've done on the computer for the past hour or so. All the bugs should—no, must—be removed from your data entry programs before they reach your investigators.

Formal Testing

Formal testing generally falls into two phases: fully automated and hands-on. The primary automated testing tool is a screen-capture utility such as AQTest, SQA Test, Silktest, and WinRunner. These utilities emulate the process a human user goes through in entering data from the keyboard, doing so at 10 times the speed, and with no lapse in attention when the same test must be repeated over and over again with minor modifications. Testing tools work in three stages:

First, they make a record of the objects—radio buttons and pop-up menus that appear on the computer screen.

Second, they record the keystrokes their users make. If the user goes to the first question and checks a "no," they record that. If he types in 11.6 in answer to the next question, they record that. When this record is played back, each and every keystroke and mouse movement of the user is repeated.

The resulting recording can be displayed by the test-program developer as a series of readily modified instructions to the computer. A standard modification consists of embedding the instructions in a loop so that the first time thorough the loop, the number 11.6 is entered, on the next occasion, 0.116, and on the next, 1160.

The reasoning behind this type of loop is that a well-designed testing program not only will use the type of data that is expected in practice but also will attempt to enter data that exceeds the preprogrammed bounds, leave some fields incomplete, and so forth.

Once such a test is developed, performing the automated test—the third stage—is as easy as pressing a button and requires an equal level of skill. A log of the test results is produced automatically and provides a permanent record of success or failure.

A further advantage of using a screen-capture utility comes when it is necessary to modify a screen by adding, deleting, or modifying any of the questions that appear on it. Once similar modifications to the testing program are completed it is ready to loop through the test again and again, a thousand times or more if necessary.

Stress Testing

Automated testing suffers from the same weaknesses as the original programming process, the inability to foresee all that a naive computer user is liable to do in practice.

You may recall the climatic scene in the film, *Good Will Hunting*, in which Robin Williams playing the part of a psychiatrist tries to persuade an anxious Matt Damon that he is not really responsible for the abuse he suffered as a child. Matt keeps saying over and over that he knows he is not responsible, but it is obvious that on a deeper level he believes quite the opposite. Most of us are the same way about computers. If we see a message that says there is a program failure, we blame it on ourselves and try to avoid all contact with that program in the future. The result of such reactions in the case of a clinical trial will be to interfere with, interrupt, and, in some instances, sabotage data collection.

The purpose of stress testing a program before releasing it for use is to detect all problems in a setting where there is little or no risk of turning off potential users. Stress testing may follow a script or may be a totally ad hoc process. A noncomputer professional should perform the test, ideally someone with a background similar to those who will be doing the data entry at the investigators' sites.

As the CRMs will be responsible for training in data entry, and must master use of the data entry screens, I recommend that the CRMs be used for the final stages of ad hoc testing so that they can combine the latter task with the former.

Warning. The project leader may need to get involved if a CRM reports that the testing is uncovering an unusually large number of errors. A meeting of the development and testing teams should be called to ensure that the project is brought back on course.

The effort preceding computer-assisted data entry is time-consuming, but it is still only a fraction of the time that will be wasted if an inferior data entry process is allowed to slip by.

TRAINING

The CRM is responsible for training all the individuals—physicians, nurses, secretaries, and technicians who will be entering the data. Thus she needs to be thoroughly familiar with the data entry process before training begins.

The training can be accomplished in either individual or group sessions. The normal sequence is to conduct training on a trial basis at one or two sites, then to give one or two group training sessions, and then to follow up with more individual sessions at sites that were missed in the group sessions along with those where groups request additional attention.

A similar strategy may need to be followed at each site with training being given initially to one or two key individuals, followed by training for all those who might have access to the computer during the trials.

I recommend that the CRM pay an initial visit to each site accompanied by the person or persons responsible for installing the computer and the data entry software. The computer and software should be brought with them rather than sent on ahead. The idea is to avoid improper or incomplete installations and to ensure that the computer is placed where it can be used conveniently during a patient examination or reading.

The training phase is never really over, as testing site personnel will continue to come and go throughout the course of a lengthy trial. Part of the monitoring process discussed in Chapter 13 consists of a review of the data entry procedures at each site.

Reminder. Training for data entry is just part of an overall training program that encompasses patient recruitment and retention. The CRM needs to sit down with the principal investigator and coordinator at each site to ensure a mutual understanding of recruitment guidelines and patient handling. Following such discussions, each site coordinator should be asked to submit a locally developed protocol for all phases of patient treatment with particular emphasis on contact and follow-up. As we discussed in Chapter 9, the objective of greater patient retention is best achieved by providing the patient with a positive reinforcing experience during each patient visit. Prolonged stays in waiting rooms or half-undressed in some isolated inside chamber do not qualify as "positive." (I hope my own physician is paying some attention to this.) Nor should the actual physician-patient contact appear rushed or hurried. VIP treatment is required for optimal patient retention. Study physicians and staff need to understand that such treatment is part of the commitment they've made.

Support

They've plopped a new computer on your desk along with software you've never seen before. You had a two-day training session (from which you had to miss four hours to take care of an emergency) and now you're on your own. Wouldn't you like to have a phone number to call just in case? A phone number where persons will be on hand to answer when you need them regardless of differences in time zones?

Now that you've seen things from the investigator's point of view, you know that a hotline needs to be part of the data entry process. Ever have the experience of being told to call back the next day because the person on the other end of the line could only answer simple questions? When you set up your hotline, staff it with knowledgeable personnel.

SUPPORT IS ESSENTIAL

Electronic clinical trials put very few technology demands on a user, and sites are usually unaware of the complexity of the underlying technology. Sites rightly expect that the technology will work when they need it and that it will not interfere with the core site functions of patient care and data collection and cleaning. While many sites have no problems running an EDC study, if technical issues develop, the sites have little or no access to the type of support they require to resolve these difficulties. Consequently, it is frequently the research staff that must take time away from their clinical tasks to work with the EDC vendor to resolve problems. The lack of local technology support and the ensuing technical demands placed on the site users are potentially serious obstacles to the acceptance of electronic clinical trials. In the long term, this problem will be alleviated as the proportion of electronic trials run at sites increases and the sites develop or outsource a local support infrastructure. EDC vendors and sponsors must carefully survey the technical support abilities of sites and, if they are insufficient, then make arrangements to either directly or indirectly provide onsite technical support.—
J. Larus, PharmaLinkFHI

BUDGETS AND EXPENDITURES

The budget for off-the-shelf hardware and software is firmed up during this phase.

I recommend the use of separate time codes to distinguish productive from nonproductive time. Delays in the arrival of hardware and software often leave programmers sitting on their hands. Similar delays arise when CRMs aren't available to answer questions. Information concerning the impact of delays is essential during post-trial review. See Chapter 15.

TO LEARN MORE

Read articles and sign up for free newsletter on issues in electronic data capture at *http://www.phaseforward.com/*.

Verweij J; Nielsen OS; Therasse P; van Oosterom AT. (1997). The use of a systemic therapy checklist improves the quality of data acquisition and recording in multicentre trials: A study of the EORTC Soft Tissue and Bone Sarcoma Group. *Eur. J. Cancer*. 33:1045–1049.

Chapter 11

Data Management

This chapter is devoted to data management, not so that you can become an expert in the area but so that you will understand the range of choices and be able to hold your own in discussions with “experts” from accounting and information systems (to say nothing of other executives who may have fallen under the spell of a salesperson). Three issues are discussed:

1. **Choice of data management software and the options available to you**
2. **Transfer of data from data entry to data storage and from data storage to your report generating and statistical analysis software**
3. **Maintaining the security and integrity of your data**

OPTIONS

Flat Files

Many managers would feel more comfortable if clinical data could be stored and viewed in a format with which they are already familiar, an Excel spreadsheet, for example (see Figure 11.1). At first glance, the spreadsheet format seems ideal: each row constitutes a different patient record, and each column a different field or variable. But as we start to fill in a mockup of our spreadsheet, two difficulties arise: first, as the number of columns exceeds the width of the screen, we may easily forget just where a particular data item is located; second, as the trial continues, we begin to accumulate multiple records for each patient—pretreatment or baseline, one-week follow-up, one-month follow-up, and so forth. Will we run out of space?

A	B	C	D	E	F
1					Recurrent Ischemia
2PATID	EVENT	PAGE	Date	Adverse Event	R2/R3
3002-1121	1MON		8/15/98	FATIGUE	
4005-1121	6MON		8/15/98	FATIGUE	
5002-1122	6MON	117_HOSP	1/31/99	EPIGASTRIC PAIN	
6002-1122	YEAR		8/26/99	UTI, HEMATURIA	
7002-1124	2WEK		9/30/98	Allergic Reaction	

FIGURE 11.1 Spreadsheet as an Example of a Flat File.

The first of these difficulties is correctable, not by Excel, but by a more-advanced flat-file manager that would allow us to search for columns by name.

The second difficulty presents more of a challenge, particularly when the different follow-ups involve different examinations and, thus, different sets of variables. While each patient's baseline record contains a host of information including demographic variables, baseline data, and laboratory values, the various follow-ups may contain only a few data items. On the other hand, the adverse event record contains many items that are not in the baseline record. When we create a column for each variable that "might" occur, the result is a worksheet made up primarily of space-consuming blank entries.

Obviously we will need several spreadsheets to store our data, perhaps one for each record type or each set of screens. But then how are we to link them in such a way that we can search and retrieve information from several worksheets at a time? Moreover, as the number and size of our worksheets grow, access times increase and corrections become more difficult.

Suppose that a follow-up exam file includes fields for the date, patient name, patient ID, patient address, plus the observations on that patient on that date. Each record must repeat the name, ID, and address of the patient increasing the amount of storage required and perhaps doubling or even tripling the time required for data retrieval.

A 10-column spreadsheet with 2000 entries requires about 200 Kbytes of storage and takes only a few seconds to sort. But a typical clinical database requires 200,000 Kbytes of storage and 1000 to 10,000 seconds to sort if the sorting methods used by Excel (one of the fastest spreadsheets) are employed.

If the patient's address changes, it will have to be changed in multiple locations or risk irresolvable inconsistencies. If the patient's name is spelled differently in different places (e.g., Phil Good, Phillip Good), then we may fail to retrieve all the necessary records.

In summary, a flat-file database like a spreadsheet contains only one record structure, many of whose fields will be empty. Access to data is done in a sequential manner; access times are slow because the entire file must be scanned to locate the desired data. Complex queries—"how many patients who were heavy smokers suffered non-Q-wave MI's during the first three months after the stent was implanted?"—are virtually impossible as there are no links among separate records.

Other problems with a flat-file database include data redundancy, the difficulty of locating and updating records as the file size increases, and the near impossibility of maintaining data integrity.[28] When the regulatory agency makes unexpected requests, will we be able to respond quickly?

Hierarchical Databases

The traditional answer to some of these issues was the hierarchical database model. A hierarchical database is a series of flat files, each one similar to a spreadsheet, that are linked in structured treelike relationships (see Figure 11.2). Data are represented as a series of parent/child relationships. A patient's record (the parent) might link to "follow-up exam" children, and each of these children might link to the records of specialized procedures (grandchildren).

Each child segment can be linked to only one parent and a child can only be reached through its parent. This could create a problem. The radiology department might want to have a patient's X-ray results as its "children." While we would want to keep them with the appropriate set of follow-ups. Or perhaps store each exam as part of a master patient record. Will we need to make two or even three copies of the exam results?

To avoid data redundancy, all information in a hierarchical database is stored in a single location and referenced by links or physical pointers in other locations. For example, the "**patients**" record might contain actual data in the "specialized exam" segment while the "**radiology**" record held only a pointer to the "specialized exam" data in "**patients**."

On the down side, the link established by the pointers is permanent and cannot be modified. A design originally optimized to work

---

[28] The lack of an audit trail would meet with fatal objections from most regulatory agencies.

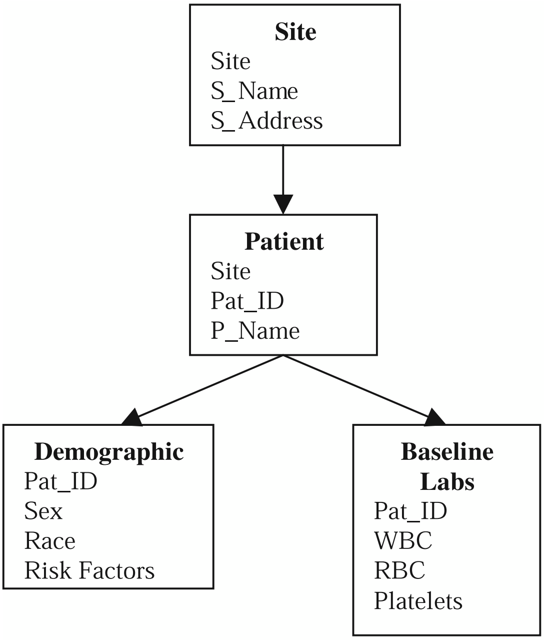

Site
Site
S_Name
S_Address
Patient
Site
Pat_ID
P_Name
Demographic
Pat_ID
Sex
Race
Risk Factors
Baseline Labs
Pat_ID
WBC
RBC
Platelets

FIGURE 11.2 Hierarchical Database.

with the data in one way may prove totally inefficient in working with the data in other ways. The physical links make it very difficult to expand or modify the database; changes typically require substantial rewriting efforts and risk introducing errors and destroying irreplaceable data.

Network Database Model

The network database model (also known as CODASYL DBTG) provides for multiple paths among segments (that is, more than one parent/child relationship) as in Figure 11.3. Unfortunately, with no restrictions on the number of relations, the database design can become overwhelmingly complex. Each new addition takes longer and longer to implement. Too often, changes that appear quick to implement at first take weeks to repair and implement correctly. The network model fails to provide the needed solution to our problems of storage and retrieval.

Relational Database Model

A relational database appears to stores all its data inside tables. Each table consists of a set of rows and columns similar (from the user's

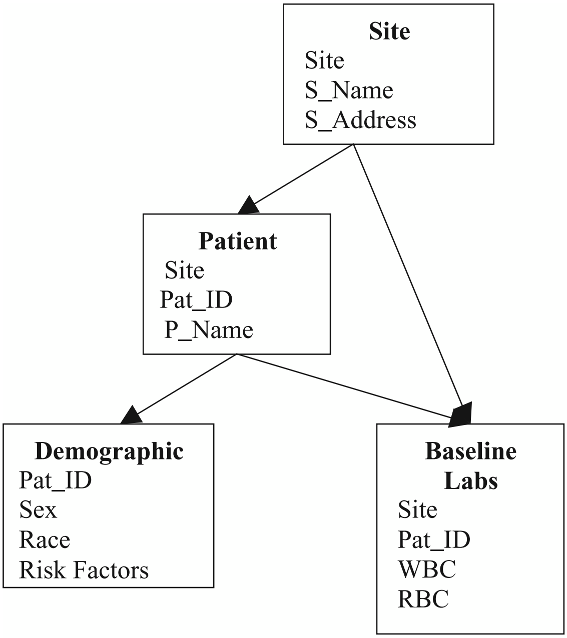

Site
Site
S_Name
S_Address
Patient
Site
Pat_ID
P_Name
Demographic
Pat_ID
Sex
Race
Risk Factors
Baseline Labs
Site
Pat_ID
WBC
RBC

FIGURE 11.3 Network Database.

point of view, though not the computer's) to the rows and columns of a spreadsheet. As with a spreadsheet, a row corresponds to a record and the columns to the data fields in the record.

Here's an example of a table and the SQL statement that creates the table:

Table: PAT_DEMOG				
Pat_ID	Sex	Brth_date	Race	Origin
071-1136	M	22/11/36	W	USA

```
CREATE TABLE Pat_Demog(
          PatID char (7),
          Sex char(1),
          Brth_date date,
                         ....)
```

In a relational database, all operations on data are done on the tables themselves, though other tables may be produced as the result. You never see anything except for tables.

Two basic operations can be performed on a relational table. The first is retrieving a subset of the columns. The second is retrieving a subset of the rows. Here are samples of the two operations:

```
SELECT PatID, Sex FROM Pat_Demog
```

Pat_ID	Sex
001-421	M
002-043	M

```
SELECT * FROM Pat_Demog WHERE sex = "F"
```

Pat_ID	Sex	Brth_Date	Diabetes
002-044	F	26/12/37	No
002-047	F	08/08/54	No

The relational approach has several major advantages including:

- **Unlimited data access**
- **Easily modified data structure**
- **Ease of access**
- **Widely used query language**

Processing queries does not require predefined access paths among the data as in a network-access database.

Changes to the database structure are easily accommodated. The structure of the database can be changed without having to change any applications that were based on that structure.

Here's an example: You add a new field for the patient's smoking habits in the patients' table. If you were using a nonrelational database, you would probably have to modify the application that will access this information by including "pointers" to the new data. With a relational database the information is immediately accessible because it is automatically related to the other data by virtue of its position in the table. All that is required to access the new field is to add it to a SQL SELECT list.

Table: **Base_Lab**				
Pat_ID	Hemo	RBC	Platelets	WBC
001-421	43	5.01	205	5
001-424	13.4	4.28	248	11.9

The structural flexibility of a relational database allows combinations of data to be retrieved that were never anticipated at the time the database was designed. In contrast, the database structure in older database models is "hard-coded" into the application; if you add new fields to a nonrelational database, any applications that access the database will have to be updated.

The true power of the relational approach comes from the ability to operate simultaneously on several tables that do *not* have the same set of columns. Suppose you want to establish a relationship between the tables Pat_Demo and Base_LAB. These tables have a common column, the name of the company. Even if each table has its own name for the column, we see that the data stored and their meaning are the same in both tables. So we could use this relationship to get a URL for each person on ADDR_BOOK. Here's the SQL statement:

```
SELECT Pat_Dem.Sex, Base_Lab.WBC
   FROM Pat_Dem, Base_Lab
   WHERE Pat_Dem.Pat_ID = Base_Lab.Pat_ID
```

This operation of matching rows from one table to another using one or more column value is called a "join," more specifically an "inner join."

A final benefit not visualized by the inventors of the relational database is that SQL, the query language associated with System R, IBM's first attempt at a relational database, has become a standardized part of all relational software. So, regardless of what brand or version of relational database your company utilizes, the same set of commands will be used to extract information. This makes it easier to transfer or bring in new employees from outside your department.[29]

Which Database Model?

The spreadsheet is ideal for organizing and presenting small amounts of data. Your staff will use spreadsheets to summarize trial findings and to communicate interim results to you.

Relational systems excel at executing complex, ad hoc queries and lend themselves to the management of large scale clinical databases.

---

[29] Visit *http://www.blnet.com/msqlpc* to get links for more information about the SQL query language.

Network and hierarchical models have a role in accounting systems where results can be reported in an unchanging straightforward format. Don't let accounting foist them off on you.

Object-Oriented Databases

Object-oriented databases can work with images, spreadsheets, documents, CAD, email messages, and directory structures, as well as text. From the computer's point of view, "data" is merely a sequence of bits (1's and 0's) residing in some sort of storage structure; perspective images are no more exotic than the characters that make up text. An object-oriented database could store ultrasound photos and ECG tracings alongside the numerical summary of results. It is difficult to visualize the practical value this still experimental approach could have for clinical trials. Let someone else be a pioneer.

CLIENTS AND SERVERS

Database management systems (DBMS), relational or otherwise, are heavily influenced by the processing architecture, that is, by how the work is divided between the host (usually a larger computer where the data are stored, a mainframe or a Unix-based minicomputer) and the client (your desktop PC or workstation). Older databases (e.g., IMS) reside entirely on the host mainframe. All Microsoft Access' data processing is done at the desktop. Somewhere in the middle is a client-server DBMS like Oracle.

I recommend the client-server approach. A database *server* is the component of a computer program that manages the database itself. The applications (queries) you and your staff work with are *clients* to the database server. In the client-server approach these applications never manipulate the database directly, they only make requests for the server to perform the operations. The server reduces the risk of data file corruption because only the server writes to the database; a crash or

WHO CHOOSES THE SOFTWARE?

Should your project be consistent with corporate guidelines? While there is a great deal of value to having corporate guidelines for software purchases including (1) reduced costs due to bulk purchase, (2) easier maintenance, and (3) ease of shifting programmers from department to department, one size seldom fits all. The key lies in knowing where to draw the boundaries.

Should clinical affairs be asked to use an inflexible decades-old data management system just because accounting does?

power outage on a client desktop—a not infrequent occurrence—will not leave the database in an incomplete state.

A database server takes advantage of the client-server architecture to lower network usage. Here is how this works: If a client application were to open a file stored on a file server directly, it would need to retrieve and transfer every record across the network just to filter out the ones it really needs. Instead, the database server filters out the unneeded records and only sends out over the network the data that really matter. Microsoft's PC-based Access® is a relational database, but it is not a database server. MS-SQL, SQL Anywhere®, DB2®, and Oracle® are both relational databases and database servers.

One Size Does Not Fit All

Bumbling's Clinical Affairs department insisted that all studies use not only the same data management system but the same database construction, the rationale being this would further reduce development costs. Here's what actually happened:

- **The cost of developing the data analysis software doubled because of the constant need to extract and merge subsets before any analysis could be performed.**
- **Because so much of each file was waste space, devoted to variables not in the current study, retrieving the information proved enormously time-consuming, and almost brought the corporate main frame to its knees.**
- **Information Systems kept buying larger and larger hard disks, but invariably ran out of space as the files with their hundreds of dummy variables were an order of magnitude larger than they should be.**
- **The complex keying instructions that resulted when several pieces of information had to be combined into one to "fit" the standard design introduced a large number of errors that had to be corrected.**
- **Programmers had to work overtime during the analysis phase designing programs that would unpack the data that had been combined and recoded during the keying process.**

COMBINING MULTIPLE DATABASES

For the reasons discussed in the preceding sections, we need to divide our data up among several files (or databases) and then to find some way of linking these files together. How should we divide the data up? A good rule of thumb is to combine all the information collected by a single individual at a single point of time in the same file.

DOMAIN TABLES

Every time you have a column whose value must come from a known set of values you have a "domain" for that column. Your domain may be as simple as the day of the week (SUN, MON, THU, . . . , SAT), as lengthy as a list of anatomical sites, or partially unknown as the set of adverse events. Domain tables reduce storage requirements, by associating an ID (or code) to a name or description. Whether the investigator enters a symptom name or its abbreviation, only one of the two (your choice) is permanently stored.

In addition to providing the benefits of flexible data entry (accepting "close" misspellings such as "asperin"), domain tables provide faster queries, faster sorting, lower storage and ease in updating. They're a standard part of most relational systems, present in both the least and the most expensive.

We link the disparate files with the aid of keys. A *key* is an element of information common to several databases that serves to tie the databases together. One obvious key is the patient ID; another is the date on which a particular examination was completed.

The key item must be stored in each of the files it is to link. A patient's name is not a good choice for a key as it is something we wish to keep confidential and, thus, stored in as few locations as possible. A patient's sex, the length of a target lesion, or the dosage of the drug the patient is receiving are not good choices since their presence in more than one database would be redundant.

A good rule of thumb is to use the patient ID as a primary key to tie all the databases together and to use an examination date or some other value as a secondary key to tie together information that is closely related. An example of related data might be the adverse event form on which the need for a certain action was recorded along with all the databases containing the information related to that action.

Further guidelines concerning keys, database structures, and data formats may be found in the documents CDISC Submission Metadata Model and Introduction to the CDISC Submissions Data Domain Models which can be downloaded from *http://www.cdisc.org/publications/index.html.*

"The Key Is the Key"[30]

Bumbling Pharmaceutical's failure to store key fields with the data cost it hundreds of thousands of dollars when its computer was unable to locate and recapture information essential to the analysis of treatment effects. Recovery was labor intensive and time-consuming, resulting in substantial delays in completion of the study.

---

[30] Whoopie Goldberg in *Jumping Jack Flash*.

The company's first error lay in adopting an older hierarchical database management system in use by their accounting staff. The rationale was that the product was already in use so the database programmers would not need to be retrained. Of course, these same programmers were already tied up with their work in accounting, so a whole new group had to be hired and trained in a system generally considered to be obsolete.

But Bumbling's second and incurable error lay in setting up the database, when the data were subdivided in accordance with the standard corporate model. Each of the case report forms was split into half a dozen parts: investigator's signature and date to one file, and the dates of various events to another. One whole file was reserved for keeping the dates of patients' visits. Unfortunately, there was no way (other than the dates themselves) to link these visits with the various follow-up forms.

An intricate coding system using a page name and an "event" would have worked had, for example, the one-month follow-up actually taken place exactly one month after the start of the intervention. But it never did except on rare occasions. Since the occasional form did not get entered in the database (and since Bumbling was using paper forms for initial data entry, there were always forms that did not get entered in the database), one could never be sure which form went with which date in the visit register.

Though not explicitly stated in the original sponsor's protocol (as it ought to have been), good medical practice requires that the taking of a repeat angiogram precede any repeats of the original procedure. The data indicating a repeat had been performed was (mainly) stored in one file whose possible keys included the patient's ID, an "event" ID, and the date of surgery. The data for the corresponding angiogram was stored in two separate files, neither of which contained the date on which the angiogram had been taken. The keys for these two files included the patient's ID, an "event" ID, and a "page" name.

A different coding system had been used for the event ID's than was used in the repeat procedure file, so it was impossible to reconcile the two.[31] Fortunately (?), there was a fourth file in which the dates of

---

[31] As with so many of Bumbling's efforts, no documentation was ever found for either coding system. Apparently, it had been left to each employee who quit, retired, or transferred to another department to brief his or her replacement.

STORE THE DATA YOU COLLECT

Although the date of discharge from the hospital was included on the case report forms, Bumbling Pharmaceuticals decided not to include this information in the database. The rationale for this deletion was lost after the database designer was promoted. Later, when it was realized the date of discharge was essential to a comparison of adverse events in the two treatment groups, the preliminary analysis had to be done by hand rather than by computer.

To avoid bumbling yourself, always

- Begin with your reports
- Plan for and collect the data you'll need for your reports
- Store the data you collect
- Store the data as they are entered
- Design the database as one file or table for each set of data entered by the same person at the same point in time

patient visits were often recorded; its possible keys included the patient ID, an "event" ID (using approximately the same coding system as the files), and the date of the visit. For repeats, the recorded visit date might or might not coincide with the date of surgery.

The only data that could be utilized for the automated (computer) analysis were results for which there was an unambiguous link between the angiogram and the repeat. As a result the angiograms for approximately 5% of the patients in the trial simply fell through the cracks. (Convert this to dollars!)

Fortunately, after a lengthy and costly inspection of the original records by hand, almost all were located. The programs used for analysis were modified to incorporate explicit reference to the individual patient records and to maintain a clear audit trail for the regulatory agency. The losses in money and time were absorbed in the company's profit and loss statement.

Transferring Data

Transferring the data entered at each physician's and each laboratory's computer to the central database can be done in one of four ways:

1. **Connect to the central computer and enter the data directly.**
2. **Enter the data into the physician's computer; transfer the data automatically at day's end via telephone to the central database.**
3. **Enter the data into the physician's computer; copy the data each day to a CD; at the end of the week, mail the CD to the sponsor to be entered into the database.**
4. **Connect to a database stored on the Web (Internet) and enter the data directly.**

TABLE 11.1 Which Data Entry Technology?

	Location of Data Entry Software	
	PC at Each Site	Web
Availability	High	Vulnerable to failures in communication lines and attacks on Internet.
Software reliability	Depends on vendor.	Relatively new technology; depends on vendor.
Ease of updating software	Must be updated separately at each site.	High
Data storage	At sponsor's site	At vendor's site
Integrity	Medium	Potential for access by unauthorized persons; Security dependent on 3rd party
Performance	High	Depends on site internet connection
Validation	Server plus site by site	Server only

Direct entry at each investigator's site to the master database is ruled out because of the impossibility of maintaining a continuous link to multiple disparate computers that may be hundreds, even thousands, of miles apart.

Which of the next two options to adopt will depend on the volume of data you expect to receive from each treatment center. With either alternative, a second data entry program "reads" the transmitted files and enters the records into the database. Human intervention is required only to start the program and slip the CD into the host computer's CD drive.

Web-based data entry is gaining increasing support. The software is the easiest to update and maintain, and one can be assured that all sites are making use of the same version. On the down side, links to the web are often down and the Web (Internet) itself isn't always up.

Separating the two stages of data entry guarantees that investigators' contact with the database is strictly limited. They can enter their data in the database, but they cannot modify it once it is entered. Nor can they access the database and be exposed to findings that might color their own observations. See Table 11.1 for a more detailed comparison.

QUALITY ASSURANCE AND SECURITY

Maintaining Patient Confidentiality

The patient's name, address, and other identifying information should be stored in one file only, and access to that file severely restricted. References to patients in all other files should be by coded ID only.

Access to Files

While the ability to *write to and modify* a clinical database can and should be restricted to a privileged and responsible (and readily monitored) few, everyone with a need to know (CRMs, project leaders, project physicians) should be able to *access and read from* the database and, moreover, to do so in a manner no more difficult than the manner in which data are entered into it.

Your data management software should permit you to establish privileges on a file-by-file basis for every individual who will have access to the system. Too often data managers and statistical analysts function as some kind of primitive priesthood, issuing proclamations that they and only they shall be privileged to access the clinical database. But with today's databases, security can be readily maintained while giving those with a need to know immediate access to the data they need.

A single individual, normally the database manager, is entrusted with issuing passwords and security levels to all those seeking access to the database.[32] Full access would include the ability to read from, write to, update, and even delete files. These privileges are necessities while the database is still under construction. Full access after the database is constructed should be limited to the database manager and his assistants.

Access should be granted on a file-by-file basis. Access to the file containing a patient's name, address, and other identifying information, and to the file containing the treatment assignments should be severely restricted. Read-only access to most of the other files should be made on a need-to-know basis. Here are two examples:

1. **The pathology laboratory should be able to both read and write to the file containing pathology reports.**
2. **Until the trials are complete, investigators should not be given direct access to the database, not even to their own data. Instead, they should submit a request to the CRM, and the CRM should provide them with any needed data that are consistent with the guidelines you and your data manager have established.**

While the database is under construction, the database manager normally makes all the decisions regarding who should have access to

---

[32] Security levels are assigned both to individuals and to automated applications.

TABLE 11.2 Tonto Project: Database Access Privileges

Name	Task	Read		Write	Create	Copy
		Baseline	Other			
Art Wood	DBM	y	y	y	y	y
Brian Donleavy	Project mgr	y	y			y
Jan Moore	CRM		y	y		y
Bill Woodson	CRM		y	y		
Mike Chuck	Statistician	y	y			
Seri Shanti	Programmer		y	y	y	

the database, the level of access they should have, and the files they should be permitted to access. Just prior to the start of full-scale data entry, the database manager should submit a chart similar to Table 11.2. to you for your approval. Thereafter, all decisions as to access to the data will be yours alone.

Maintaining an Audit Trail

To avoid even the appearance of fraud, and to satisfy government regulations, all changes to the database must be automatically tracked and recorded by the system.

Security

The best approach to security is pure paranoia, although your chief concern is not theft but the integrity and safety of the database. Access by your competitors is seldom a problem. The very fact you are conducting full-scale clinical trials tells them most of what they want to know. Other information they will attempt to glean from your employees.[33]

To protect the database, you will need to backup (copy) and test your data on a regular basis and store at least some of the backup copies in a remote location. Standard practice is the cycle of 3 (sometimes 5, and sometimes 7):

Backups are made on the evenings of days 1, 2, and 3. On the night of day 4, the backup of day 4 is sent to a remote location, and the cycle is restarted; the backups of days 2 and 3 are discarded as days 5 and 6 are backed up.

While backing up the database and storing the backups at a remote location is essential, you'll need to do more. Backing up data that are already contaminated is pointless. You'll need to run tests on your database before a backup is made. These tests are particularly important during the early part of the trials when you are unlikely to be accessing the data on a regular basis.

---

[33] To forestall leaks, treat your employees' as you would be treated. See Gandy (2001) and Mendes (1995).

Tests should be made on a file-by-file basis. Possible tests include:

- **Selecting five records at random from each file**
- **Counting the total number of records**
- **Printing out the minimum and maximum of at least two of the fields**

FOR MORE INFORMATION

Date CJ. (1999). *An Introduction to Database Systems* 7th ed. San Francisco: Addison-Wesley.

Garcia-Molina H; Ullman JD; Widom J. (1999). *Database System Implementation*. Upper Saddle River, NJ: Prentice Hall.

Gandy B.G. (2001). *30 Days to a Happy Employee : How a Simple Program of Acknowledgment Can Build Trust and Loyalty at Work*. New York: Fireside Press.

Mendes A. (1995). *Inspiring Commitment: How to Win Employee Loyalty in Chaotic Times*. New York: McGraw-Hill.

Prokscha S. (1999). *Practical Guide to Clinical Data Management*. Englewood, CO: Interpharm.

Rondel RK; Varley SA; Webb C. (2000). *Clinical Data Management*, 2nd ed. New York: Wiley.

Chapter 12

Are You Ready?

THIS CHAPTER IS INTENDED to serve as a master checklist before you start the actual study. Most of the points touched on here, many of them quite minor, have already been disposed of by your staff. But as with a vacation whose first few days are ruined because you forgot the mosquito repellent, a few extra moments of reflection before your departure can yield large dividends.

We begin by reviewing the basis of your study, the pharmaceuticals or devices that will be used in the intervention—you do have a full supply on hand!—and then go over points to be covered with your investigators and site coordinators, the final field tests of your questionnaires, data entry software and hardware, and uniform instructions to be issued to participants.

PHARMACEUTICALS/DEVICES

- **Preferably all devices and drugs used in the study should be drawn from the same lot and set aside before the start of the study. Obvious exceptions include prohibitively expensive devices that are normally manufactured on a just-in-time-basis and pharmacological agents whose potency deteriorates quickly.**
- **Controls should be matched to the active intervention on the basis of both appearance (size and color) and taste. The matching should be verified by your staff.**
- **All vials should be labeled with the patient's name, the patient's ID, and (if applicable) instructions for self-administration (e.g., twice a day with meals).**
- **In order to monitor compliance, the patient should only be supplied with the amounts of pharmacologic agents needed for use until the next scheduled checkup.**

- **The initial labeled supplies should have been shipped to the investigators.**

SOFTWARE

- **Data entry software is completely debugged and tested.**
- **The variables and layouts of the interim reports have been determined.**
- **Software for tracking recruitment is in place and tested.**

HARDWARE

- **All the computer hardware to be provided to investigators has been purchased and delivered.**
- **All units were equipped with software and tested in house before shipping them to investigators.**
- **All units were installed and field tested at the individual investigators' sites.**
- **Service personnel whose primary responsibility is to maintain the equipment in the field have been hired and trained.**

DOCUMENTATION

The following are complete and submitted or printed:

- **Proposal**
- **Physician's manual**
- **Informed consent form**
- **Laboratory manuals**
- **Patient instructions**

INVESTIGATORS

Investigators have been recruited. Your staff has visited the sites of all investigators and investigational laboratories. At each site you have done the following:

- **Established the identity of the single individual who will be responsible for ensuring the orderly collection and monitoring of data.**
- **Conducted training programs for the chief investigator and all individuals who will have contact with patients or be involved in data entry.**

- **Provided copies of procedure manuals and of the forms to be given to patients. These latter include informed consent, study overview, and any instructions needed to ensure compliance.**
- **Supervised the installation of the data entry computer and field tested the software, the hardware, and the communication links with your office.**

REVIEW COMMITTEES

- **Referral criteria determined. (Will the pathology committee review all biopsies or only those biopsies that are suspect or ambiguous?)**
- **Duties determined. (For example, the safety committee will classify all adverse events as to whether or not they are treatment related.)**
- **Meeting arrangements determined. (For example, the members of the safety committee will each make an independent review, then teleconference monthly for a final determination.)**
- **All committee members recruited.**

PATIENTS

- **Recruiting quotas determined for each treatment site.**
- **Recruiting methods determined subject to ongoing review.**
- **Patient instructions ready for field test.**

REGULATORY AGENCY

- **Agency has approved the trials.**

TEST PHASE

In most cases it is advantageous to hold a preliminary test phase at one or two sites to ensure that all systems are fully functional. A mixture of site types and logistical support capabilities is essential. The evaluation would include field tests and possible revision of

- **Drug/device delivery procedures**
- **Data entry software**
- **Transfer of data from the investigator to central storage**
- **Physician and laboratory manuals**

- **Recruitment procedures**
- **Patient instructions**

A report on these tests with a list of proposed modifications should be made to the implementation committee.

Chapter 13

Monitoring the Trials

AS WITH EACH OF THE MANY PROJECTS WITH which you as a manager will be associated, success can only be assured if you plan exhaustively (the focus of the last eight chapters), carefully monitor the project's progress (the topic of the next three), and review the results to see what may be gained from your experiences for use in the future.

Ongoing monitoring is essential for all of the following:

- **Recruiting and retaining investigators**
- **Recruiting and retaining patients**
- **Protocol adherence**
- **Quality control**
- **Patient compliance**
- **Limiting adverse effects**

ROLES OF THE MONITORS

Most of the duties, if not all of the responsibility, for monitoring the trials will fall upon the shoulders of the medical monitor and the clinical research monitor(s), or CRMs.

The medical monitor is responsible for:

1. **Preliminary site visits at which the investigator's bona fides are established.**
2. **Monitoring the progress of physician recruiting.**
3. **Monitoring all remaining aspects of the trials including:**
    a. **Patient withdrawals and noncompliance**
    b. **Adverse events**
    c. **Other problems brought to her attention by the CRMs**
4. **Taking such corrective action as may be deemed appropriate including:**

        a. **Arranging for review by an independent pathologist or radiologist**
        b. **Bringing protocol violations, excessive numbers of adverse events, and treatment-related dropouts to the attention of internal and external review committees**

5. **Maintaining physician interest throughout lengthy trials**

The medical monitor should write out the responsibilities of the clinical investigators, provide training in these responsibilities, remind investigators of their responsibilities as necessary, and with the assistance of the CRMs reward investigators for successful execution of their responsibilities.

The CRM is responsible for:

- **Preliminary visits to each site to ensure the smooth flow of data.**
- **Monitoring the progress of patient recruiting.**
- **Working with the statistician to develop interim reports, and reviewing these reports on a regular basis.**
- **Monitoring and correcting problems at the investigator's site including those related to:**
    - **— Drug/device delivery**
    - **— Data entry**
    - **— Transfer of data from the investigator to central storage**
    - **— Procedures for treatment and observation**
    - **— Availability of manuals and forms**
- **Maintaining physician interest throughout lengthy trials.**
- **Follow-up site visits that serve both to maintain morale and to forestall problems with patients, the investigator, and the investigator's staff.**
- **Picking up and verifying the signed, printed case report forms and seeing that the investigator is paid in timely fashion. (We strongly recommend against separating the accounting function from that of the monitor.)**
- **Referring otherwise intractable problems to the medical monitor or project manager.**

CRMS PRETRIAL CHECKLIST

Establish a delivery schedule for drugs/devices.

At each investigator's site ensure that:

  Initial supplies of drugs/devices are present.

  Data entry hardware and software is installed and tested.

  Training is completed.

  Site coordinator is employed/appointed.

  Informed consent forms are on hand.

  Procedure manuals are on hand.

  Instructions, containers, and labels for shipping to investigational laboratories are on hand.

The level of monitoring will depend on the nature of the trial (Phase I, II, or III) and its intent. Monitors need to be aware of patient withdrawals and to discuss them either with the planning committee (if the withdrawals affect all sites) or with the specific investigators involved.

BEFORE THE TRIALS BEGIN

Before the trials begin your CRMs and medical monitor should already have met not once but several times with each investigator. The purpose of these meetings is threefold:

1. **To survey the investigator's working arrangements.**
2. **To establish rapport with and gain the trust of the investigator.**
3. **To arrange for the appointment of a site coordinator.**

Establishing rapport is often quite difficult as the responsibilities borne by investigators tend to make them suspicious of all those without similar training. Physicians trust physicians, pilots trust pilots, and cops trust cops. Everyone else is a "civilian." Even the appointment of a site coordinator often meets with some resistance. Experienced monitors know that investigators tend to fall into one of three categories:

1. **The "professional" understands both the need for and the value of a site coordinator and is eager to turn over the responsibilities.**
2. **The "do-it-myselfer" actively resists the appointment.**
3. **The "amateur" appears to go along with the request but actually sabotages it, appointing a temporary employee or a nurse who is about to depart on maternity leave.**

Each category requires quite different treatment.

The experienced monitor understands the value of the professional and does not impose upon her time. Direct contacts with the professional—phone calls and visits—are kept to a minimum. Only when there are problems that cannot be resolved by the site coordinator is direct contact made.

The professional should be provided with regular progress reports concerning the study as a whole. "Rewards" might take the form of tickets to some event in which the investigator has expressed an interest.

The "do-it-myselfer" requires additional reassurance before being willing to delegate authority. She should be reminded that invariably she will be called away at a critical moment and, in any event, will need to designate a subordinate to take over her duties when she is summoned to an investigator's meeting. (You needn't actually have scheduled an investigator's meeting to utilize this ploy.) If all else fails, the monitor should simply say that company policy requires that a site coordinator be appointed.

The project manager and medical monitor must back up the CRM on this point. It is essential to the validity of the study that the CRM meet with every member of the investigator's staff who will have any contact with patients in the study. Absent such contact, the investigator cannot be allowed to participate.

Monitors should make frequent contact with this type of investigator both prior to and during the study. To forestall problems, experienced CRM makes it a point to notify the investigator prior to any contact with the site coordinator. This type of investigator should be provided with frequent progress reports on the conduct of the study.

The "amateur" also requires extended contact until the need for a full-time study coordinator working under the authority of the investigator is understood. One possible motive for a lack of cooperation is the need of the investigator to retain her staff for other purposes. Be prepared to either pay for an on-site coordinator or forgo the investigator's services. This type of investigator also requires frequent monitoring initially.

KICK-OFF MEETINGS

It is almost a tradition in the pharmaceutical industry that each set of clinical trials be preceded by a kick-off meeting at which the investigators are wined and dined at some exclusive retreat. We see no value in this. Indeed, such meetings often have a negative impact with various groups of investigators reaching independent informal agreements on how they will conduct their portion of the study.

It is far better to reserve discretionary funds for morale-building meetings midway through or near the close of multiyear trials. Such follow-up meetings can also be used to go over aspects of the trials that have shown wide divergence among investigators.

DUTIES DURING TRIAL

Physician liaison is easily a full-time job. (If a study takes five years, the full-time job will last all five.) Most common problems can be solved by effective communication between the monitors and the investigator. The efforts of the monitors may need to be supplemented by those of programmers, document writers, field engineers, the statistician, and virtually anyone who participated in the development phase.

Site Visits

Site visits need to be disciplined and organized with the organization taking place before the monitor leaves his or her office. This includes reviewing all previously submitted forms to see if they have been completed unambiguously. (See also under the heading "quality control" in the next chapter.) Too frequent use of "other and unspecified" categories should be a red flag. The frequencies of adverse events, missing forms, and patient withdrawal should be compared with those at other study sites.

I recommend each visit include a sit down meeting with the investigator and/or the site coordinator whether or not there are problems. The purpose of such a meeting is primarily motivational, to share with the investigator information concerning the overall status of the project. Of ongoing concern is the priority the investigator is giving to your clinical trial—it cannot be allowed to drop off his or her viewing screen. Additional compensation should be provided (tickets to movies or sports events, or a restaurant gift coupon) on an ongoing basis to the individual(s) responsible for data entry at each site.

Visits should rarely be made on a fixed frequency to a given location. A site should be visited more frequently and the stay be of longer duration if the site is new to the study or has a prior history of problems.

At least one visit should be made simply to ensure that

- **Assignment to treatment is adhered to—in inadequately monitored single-blind studies, physicians have been known to ignore the treatment assignment from day 1.**
- **Treatment procedures are adhered to.**
- **Data recording methods are adhered to.**
- **Informed consent forms are administered correctly.**
- **Samples and specimens requiring off-site review are dispatched promptly.**

Sites with novice investigators, lagging enrollment, delays in submission of informed consent and other forms, and or exceptional frequencies of adverse events should be visited more frequently.

The project manager needs to oversee the monitors' schedules; personality conflicts are not unknown and there is a natural reluctance on anyone's part to return to an area where a hostile reception is anticipated.

Between Visits

The CRMs should remain in contact with investigators and supporting laboratories between visits so as to be aware of any or all of the following:

- **Staff turnover**
- **Changes in investigator's other responsibilities**
- **Changes in facilities**
- **Patient deaths or severe adverse reactions**
- **Protocol deviations**
- **Cost overruns**
- **Loss of interest**

Staff Turnover. In the event of excessive staff turnover, be prepared to provide supporting staff and to train new personnel.

Changes in Responsibilities. One reason for ongoing contact with investigators and their staff is try to head off diversions of the investigator's time. In that event you may have to monitor the investigator's activities more closely and revise downward your estimate of the number of patients likely to be treated at that site.

Changes in Facilities. Changes in facilities range from construction that forces patients to wait in the hall to out-of-alignment measuring devices to data entry terminals that have been unplugged. Working together, the CRM and the site coordinator can often resolve many of the resulting problems. For example, while screaming at a hospital administrator will not generate more space in an already crowded institution, direct contact with study subjects who have been inconvenienced can be used to motivate them to remain with the study.

Occasionally, a monitor is confronted with a series of handwritten data entry forms and is told that the computer has been down for a week or two. Too late to ask, "Why wasn't I informed?" The solution lies with training, with encouraging the investigator's staff to contact your staff if there are further problems and, not least, with following through on such contacts.

Patient Deaths or Severe Reactions. In the event of patient deaths or severe reactions, the investigator will often want to crack the treatment code. She should be discouraged from doing so until the medical monitor has had the opportunity for a thorough review of the case.

Deviations from Protocol. Deviations from protocol including modification of treatment without notification cannot be allowed to go undetected or unremarked. Immediate face-to-face contact by the medical monitor with the offending investigator is called for. Listening is always more effective than telling, but, inevitably, agreement to future adherence must be reached or the investigator's services terminated.

Loss of Interest. Loss of interest is not infrequent in long-term studies. A variety of methods for handling the problem are considered in the final section of this chapter. I strongly urge the production and distribution of a newsletter reporting on the trials at between two- and three-month intervals as an inexpensive method for both maintaining interest and disseminating information.

Cost Overruns. Depending on the type of billing arrangements, either you or the investigator may first realize that expenses at the investigator's site are excessive. The CRM's responsibility is to document all the expenses involved and be in a position to compare them with expenses at other sites. If it is the investigator who feels

NEVER GIVE AN ADVANCE—OR HOW 300 PATIENTS TURNED INTO 75

Several years ago I was called in as a consultant by a medical-device firm, ostensibly to do a survival analysis requested by a federal agency, but in actuality (the almost invariable case when a consultant is brought in), to be the bearer of bad news concerning the many failed aspects of the study.

The data management system crashed regularly. The database was filled with discrepancies, and there were no procedures in place for remedying them. As I proceeded step by step with the analysis, the 300 case report forms that had been filed so promisingly at the beginning of the study were reduced to 75 when I looked at the follow-ups. Five of the missing forms could be found on the CRM's desk—the rest simply had never been submitted (and, in some instances, never completed) by the physician.

You'd think the cardiologists would have wanted their money. They'd done the work; all they had to do was return the forms and the dollars would follow.

"They've already been paid," was the unexpected response to my query. "Most are friends of the president of our company, he's a physician, too, and he made sure they were paid promptly."

"Without checking to see if they'd done the work?"

"Without checking."

As it turns out, this company's device did increase the survival rate of the cardiac patients who were fortunate enough to receive it. But it took the company another year before sufficient data were on hand to validate their claim, and another year of deferred profits (and wasted lives) before their product could be brought to market.

aggrieved, the CRM (and the medical monitor) must listen and respond to the investigator's feelings.

The possible actions you as project manager might take are discussed under budgeting in the next chapter. Because of the monitors' continuing need to maintain rapport with investigators, I recommend that a third party be the one dispatched to work with the investigator if corrective action is called for.

Other Duties

Other duties of the monitors during the trials include

- **Arranging for payment of physicians and testing laboratories as completed reports are received (see sidebar).**
- **Monitoring adverse effects, a topic we consider at length in the following chapter.**
- **Maintaining physician interest in lengthy trials.**
- **Revising interim reports. Although sketches of the form and layout of the interim reports will have been developed prior to the trials, the details invariably evolve as the trials progress. Frequencies are replaced or supplemented by percentages, graphs supplement text, and additional breakdowns by patient and disease characteristics required. The CRM and medical monitor may also need to request comprehensive workups for specific sites that the monitor intends to visit.**

MAINTAINING PHYSICIAN INTEREST IN LENGTHY TRIALS

With patients being recruited into the study over a long period of time, participating physicians may lose interest when they fail to see dramatic results in patients who are further along in the process. Look to physician motivations to ensure retention. Often what was a motivating factor at the start of the trials proves less rewarding than hoped for or even turns into an irritant.

Let's take a second look at the list of potential motivators that was provided in the chapter on physician recruiting:

1. **Enhance one's career—but only if the trial is paying off.**
2. **Participate in scientifically exciting research—continual feedback on the study's progress is called for.**
3. **Obtain medical benefits for one's patients—again, you may need to remind the physician of the nature of these benefits; a brief irregular newsletter, distributed in both hard copy and email form can prove valuable.**
4. **Obtain new medical or scientific equipment provided by sponsors to enable trial (or purchased with monies only available down-**

stream)—if already purchased the investigator may have come to view such equipment as his own; three or four months into the trials would be the ideal time to affix a plaque with the trial Sponsor's name to each piece of donated equipment.

5. **Publish scientifically and medically important article—again, this is only a motivator if it appears there will be positive publishable results.**
6. **Obtain new staff to help with clinical trial—your trials may come to be viewed as a burden as the physician begins to think of the staff as her own; summoning all sponsor-paid staff to your offices for a one-day meeting can provide an essential reminder.**

Some investigator motivators remain of continued importance, provided that investigators do not forget who their source is:

7. **Obtain money that may be used for personal interests.**
8. **Obtain money that may be used to conduct unsponsored trails of professional interest.**
9. **Enhance one's standing in an academic department because of bringing in grant money.**
10. **Develop a long-term relationship with a sponsor.**
11. **Repay a favor (can only be pushed so far).**

With these latter motivators, you may need to adopt a carrot and stick approach, withholding monies or offering additional inducements. As always, the most cost-effective approach is to make the investigator feel part of a larger, inherently rewarding group experience.

Chapter 14

Managing the Trials

IN THE PREVIOUS CHAPTER WE CONSIDERED those activities for which the CRM and medical monitor bear primary responsibility. In this chapter we consider those functions where the CRM's role is secondary, though equally essential, as a highly skilled triage coordinator and data gatherer.

The day may come when some advanced AI-equipped cyborg will take over the CRM's monitoring activities; until then we must rely on humans to go over the interim reports and take whatever actions are warranted.

Examining a report on recruitment, for example, the CRM must determine whether the problem can be localized to a few sites or is more widespread. She must decide whether the site merely requires moral encouragement or whether more active intervention is warranted. Should the medical monitor be notified and requested to intervene? Or can the CRM handle the problem alone?

Similarly the CRM must decide whether to take direct action in problems involving withdrawals, missing data, or requests from investigators for additional information or guidance, or to refer the problems and requests to other individuals.

Problems may arise with any or all of the following:

- **Recruitment**
- **Late and incomplete forms**
- **Dropouts and withdrawals**
- **Protocol violations**
- **Adverse events**

RECRUITMENT

The statistician can play an invaluable role in this function by developing forecasting models for recruitment. In those instances where recruitment is essentially passive—with prospective subjects appearing at investigators' offices purely as a matter of chance—forecasting is relatively straightforward. In Figure 14.1, a simple linear regression model provides an appropriate forecast.

The response to active recruiting tends to take the form shown in Figure 14.2 of a sharp rise followed by a relatively slow drop-off in numbers. The trick to forecasting is realizing when the point of diminishing returns is reached and further investment in recruiting would be a waste of time.

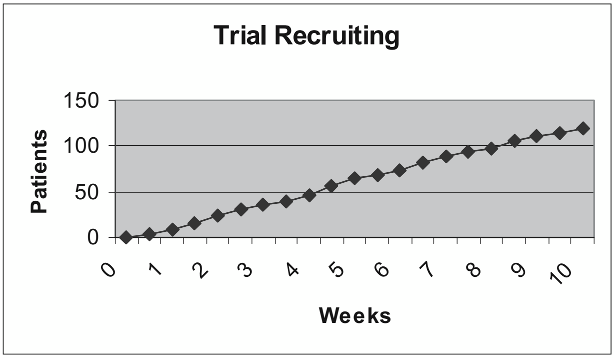


FIGURE 14.1 Forecasting Patient Recruitment

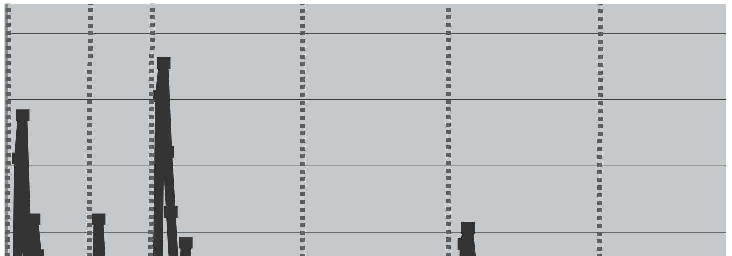


FIGURE 14.2 Response to active recruiting. Each peak represents a response to a specific recruiting campaign. Note the reduced response to later campaigns.

Late and Incomplete Forms

Late and incomplete forms are the major source of cost overruns and trial delays. They cannot be tolerated. Automated transfer of data simplifies the preparation of a summary report detailing the forms received and the date of the next form due by site and patient ID. This report should be updated and reviewed daily by the CRM. Overdue and incomplete forms should be flagged as in Table 14.1.

The standard procedure for dealing with delays is as follows:

1. **The CRM places a telephone call to the site coordinator to determine the source of the difficulty.**
2. **She does what she can to facilitate collection and transmission of the needed information.**
3. **If the missing data involve several patients at the same site, she may choose to visit the site or to refer the matter to the medical monitor.**
4. **In turn, the medical monitor may either deal with the problem(s) or refer them to the project manager.**
5. **The primary responsibility of the project manager is to ensure that procedures are in place and that decisions are made not deferred.**

Dropouts and Withdrawals

Missing or delayed forms are your first indication of problems involving dropouts or withdrawals. The first step is to determine whether the problem can be localized to one or two sites. If the problems are widespread, they should be referred to the biostatistician (who has access to the treatment code) to determine if the withdrawals are treatment related.

Problems that can be localized to a few sites are best dealt with by a visit to that site. Widespread problems should be referred to an internal committee to determine the action to be taken.

Dropouts and withdrawals should be analyzed to see if they are treatment related. (The patient feels better and decides "why bother," or feels worse and draws the same conclusion.) In such circumstances,

TABLE 14.1 CRFs from Examining Physician—Oct 10 1999

Site	Patient	Elig	Baseline	2 wk	1 mo	2 mo	3 mo	6 mo	1 yr
001	100	6/3/99	6/18/99	7/8/99	7/22/99	8/25/99	**9/25/99**		
001	101	7/8/99	8/01/99	8/15/99	8/31/99	9/30/99	10/30/99		

missed or reduced doses also are to be suspected. A subsample of the patients who have dropped out should be questioned, an internal committee review the data, and appropriate action taken.[34]

Protocol Violations

Suspected protocol violations should be referred to the medical monitor for immediate follow-up action.

As discussed in Chapter 7, a variety of corrective actions are possible from revising the procedures manual if its ambiguity is the source of the problem to severing ties with a recalcitrant investigator. The CRM is responsible for recording the action taken, and continues to be responsible for monitoring the out-of-compliance site.

CLINICAL TRIALS REPRESENT A LONG-TERM COMMITMENT

Bumbling lost interest in their Brethren device test midway through when they realized the results just weren't going to come out the way they planned. They probably would have shelved the project indefinitely had not it been brought forcefully to their attention that when you experiment with human subjects, the government insists on knowing the results whether or not they favor your product.

Adverse Events

Excessive numbers of serious adverse events can result in decisions to modify, terminate, or extend trials in progress. Comments from investigators along with ongoing monitoring of events will provide the first indications of potential trouble.

Comments from investigators commonly concern either observed "cures" (generally acute) or unexpected increases in adverse events. Both are often attributed by investigators to the experimental treatment, even though in a double-blind study the code has not yet been broken.

The CRM, working in conjunction with the medical monitor, compares the actual numbers with the expected frequency of events in the control or standard group. If the increase appears to be of clinical significance, the statistician is asked to provide a further breakdown by treatment code.

While the statistician will report the overall results of her analysis, neither the CRM nor the medical monitor, both of whom work directly with the investigators, should come in direct contact with the uncoded data. For the same reason only aggregate, and not site-by-site, results should be reported.

If the results of the analysis are not significant or of only marginal statistical significance (e.g., 5% level), the trials should be allowed to continue uninterrupted.

---

[34] Any modifications to treatment will require that the regulatory agency be notified.

If the results are highly significant, suggesting either that the new treatment has a distinct advantage over the old or that it is inherently dangerous, a meeting of the external review committee should be called.

QUALITY CONTROL

Quality control is an ongoing process. It begins with the development of unambiguous questionnaires and procedure manuals. And ends only with a final analysis of the collected data.

Interim quality control has four aspects:

1. **Ensuring the protocol is adhered to, a topic discussed in the previous chapter.**
2. **Detecting discrepancies between the printed or written record and what was recorded in the database, a problem minimized by the use of electronic data capture.**
3. **Detecting erroneous or suspect observations.**
4. **Putting procedures in place to improve future quality.**

The use of computer-assisted direct data entry has eliminated most discrepancies of the first type with the possible exceptions of the results of specialty laboratories, which are used so infrequently that supplying them with computers would not have been cost effective, and the findings of external committees, which are normally provided in letter form.

Confirmation and validation of specialty laboratory results is normally done in person, perhaps no more often than once every three months.

The findings of external committees often arrive well after the other results are in hand. They are often transcribed and kept in spreadsheet form. While such spreadsheets can be used as a basis for analysis, I'd recommend that they be entered into the database as soon as possible. Here's why: The spreadsheet is often too convenient, with the result that multiple copies soon are made, each copy differing subtly from the next with none ever really being the master. A single location for the data makes it easier to validate each and every record against the original printed findings of the external committee.

One quick way to detect suspect observations, particularly for calculated fields, is to prepare a frequency diagram. In Figure 14.3, prepared with Stata©, a set of ultra high observations well separated

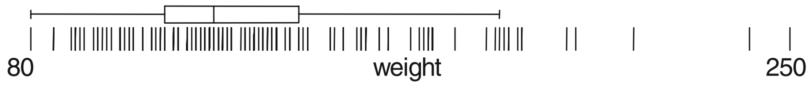


FIGURE 14.3 Box and Whiskers Plot. Display of weights of 187 young adolescent female patients with a box and whiskers plot superimposed above. The two largest values of 241 and 250 pounds seem suspicious. Better double check the case report forms.

from the main curve stands out from the rest. Sorting the data quickly reveals the source of the suspect values; the SAS Univariate procedure, for example, automatically tabulates and displays the three largest and smallest values.

ROLES OF THE COMMITTEES

Recall that external committees have three basic functions:

1. **Interpretation of measurements. Does the ECG reveal an irregular heartbeat?**
2. **Assigning causes for adverse events. Was the heart attack related to treatment?**
3. **Advising on all decisions related to modifying, terminating, or extending trials in progress.**

We consider the functions of the first two types of committee in this section and of the latter, trial-review and safety committee in the following section.

The initial meeting of each committee should be called by the medical monitor. Procedures for resolving conflicts among committee members (rule by majority, or rule by consensus with secondary and tertiary review until consensus is reached) should be established.

After the initial meeting, members of these committees no longer need, in theory, to meet face to face. At issue is whether decisions should be made independently in the privacy of their offices or at group sessions. This problem is an organizational one. Will less time be spent in contacting members one by one (the tardy as well as the prompt) to determine their findings? Or in delaying meeting until a group session can be scheduled?

The chief problems related to these committees have to do with the dissemination of observations to committee members, the collection of results, and the entry of results into the computer.

Today digital dissemination on a member-by-member basis is to be preferred to the traditional group meeting. Problems will arise only if a committee member lacks a receiving apparatus. It is common to use the same individuals on multiple studies thus justifying the purchase of such equipment for them.

Members should be given a date for return of their analysis. The CRM should maintain a log of these dates, following up with immediate reminders should a date pass without receipt of the required information.

The CRM should maintain a spreadsheet on which to record finding from committee members as they are received. Spreadsheet data may then be easily entered into the database by direct electronic conversion.

Committee members require the same sort of procedure manuals and the same sort of follow-ups as investigators.

TERMINATION AND EXTENSION

Several stages and many individuals are involved in decisions to modify, terminate, or extend trials in progress. In this section we detail the procedures and decisions involved.

A meeting of the external safety review committee should be called if either there have been an excessive number of adverse events or a medically significant difference between treatments has become evident.

The statistician should prepare a complete workup of all the findings as she would for a final report. The medical monitor should convey the findings to the external review committee. The CRMs and the statistician should accompany him in case the committee has questions for them.

The safety committee has two options:

1. **To recommend termination of the trials because of the adverse effects of the new treatment**
2. **To recommend modification of the trials**

Such modification normally takes the form of an unbalanced design in which a greater proportion of individuals are randomized to the more favorable treatment.[35]

In such an *adaptive design*, the overall risk to the patients is reduced without compromising the integrity of the trials. The only "cost" is several more days of the statistician's time and several minutes of the computer's.

At issue in some instances is whether individuals who are already receiving treatment should be reassigned to the alternative treatment. Any such decision would have to be made with the approval of the regulatory agency.

---

[35] See, for example, Armitage (1985), Lachin et al. (1988), Wei et al. (1990), and Ivanaova and Rosenberger (2000).

In any event, observations on individuals already enrolled should continue to be made until the original date set for termination. This is because a major purpose of virtally all clinical trials is to investigate the degree of chronic toxicity, if any, that accompanies a novel therapy. Notably absent from our list of alternatives is the decision to terminate the trials at an early stage because of the demonstrable improvement provided by the new treatment.

EXTENDING THE TRIALS

After a predetermined number of individuals have completed treatment, but before enrollment ceases, the project manager should authorize the breaking of the code by the statistician and the completion of a preliminary final analysis. As previously noted, the statistician should be the only one with access to the decoded data, and the results should be reported on an aggregate, not a site-by-site basis.

A WORD OF CAUTION OF SPECIAL INTEREST TO CUBS FANS

Though tempting, decoded results, broken down by treatment, should *not* be monitored on a continuous basis. As any stock broker or Cubs fan will tell you, short-term results are no guarantee of long-term success.

In July 2001, baseball's Chicago Cubs were in the lead once again, a full six games ahead of their nearest inter-division opponent. Sammy Sosa, their right fielder, seemed set to break new records.[36] Moreover the Cubs had just succeeded in acquiring one of the Major League baseball's most reliable hitters. Success seemed guaranteed.

Considering that the last time the Cubs won the overall baseball championship was in 1906, a 20-game lead might have been better. The Cubs completed the 2001 season completely out of the running.

Statistical significance early in clinical trials when results depends on only a small number of patients offers no guarantee that the final result will be statistically significant as well. A series of such statistical tests taken a month or so apart is no more reliable. In fact, when repeated tests are made using the same data, the standard single-test *p*-values are no longer meaningful.

Sequential tests, where the decisions whether to stop or continue are made on a periodic basis, are possible but require quite complex statistical methods for their interpretation.[37]

---

[36] He later broke several records.

[37] See, for example, Slud and Wei (1982), DeMets and Lan (1984), Siegmund (1985), and Mehta et al. (1994).

If significant differences among treatment groups are observed, then the results may be submitted to an external committee for review. If the original termination date is only a few weeks away, then the trials should be allowed to proceed to completion.

If the differences among treatments are only of borderline significance, the question arises as to whether the trials should be extended in order to reach a definitive conclusion. Weighing in favor of such a decision would be if several endpoints rather than just one point in the desired direction.[38] Again, the matter should be referred to the external committee for a decision, and if an extension is favored by the committee, permission to extend the trials should be requested from the regulatory agency.

BUDGETS AND EXPENDITURES

I cannot stress sufficiently the importance of keeping a budget and making an accounting of all costs incurred during the project. This information will prove essential when you begin to plan for future endeavors.

Obvious expenditures include fees to investigators, travel monies, and the cost of computer hardware and over-the-counter software. Time is an expenditure. Since most of us, yourself included, will be working on multiple projects during the trials, a time sheet should be required of each employee, and a group of project numbers assigned to each project.

Relate the work hours invested to each phase of the project. Track the small stuff including time spent on the telephone. The time recorded can exceed 8 hours a day and 40 hours a week, and often does during critical phases of a clinical trial. (These worksheets also provide a basis for arguing that additional personnel are required.)

A category called "waiting-for" is essential. With luck—see the final chapter—we can avoid such delays the next time around. Also of particular importance in tracking are tasks that require time-consuming manual intervention such as reconciling entries in "other" classifications, and clarifying ambiguous instructions.

Midway through the project you should be in a position to finalize the budget. Major fixed costs will already have been allocated. And the average cost per patient determined.

If you've followed the advice given here, then even the programming required for the final analysis should be 99% complete—and so too will be the time required for the analysis. While developing programs for statistical analysis is a matter of days or weeks, executing the completed programs against an updated or final database takes only a few minutes. Interpretation may take a work-week or more with several additional work-weeks for the preparation of reports.

---

[38] A multivariate statistical analysis may be appropriate; see Pesairn (2001).

Ours is a front-loaded solution. Savings over past projects should begin to be realized at the point of three-quarters completion, with the comparative numbers looking better and better with each passing day.

If you've only just adopted the use of electronic data capture, there may or may not be a record of past projects against which the savings can be assessed. The costs of "rescue efforts" often get buried or are simply not recorded. Thus the true extent of your savings may never be known. All the more reason for adopting the Plan–Do–Check approach in your future endeavors. Undoubtedly, changes in technology will yield further savings.

TO LEARN MORE

Armitage P. (1985). The search for optimality in clinical trials. *Int. Stat. Rev.* 53:15–24.

DeMets DL; Lan G. (1984). An overview of sequential methods and their application in clinical trials. *Commun. Stat. Theory Meth.* 13:2315–2338.

Fleming T; DeMets DL. (1993). Monitoring of clinical trials: issues and recommendations. *Control. Clin. Trials* 14:183–197.

Hamrell MR, ed. (2000). *The Clinical Audit in Pharmaceutical Development*. New York: Dekker.

Ivanova A; Rosenberger WF. (2000). A comparison of urn designs for randomized clinical trials of $K > 2$ treatments. *J. Biopharm. Stat.* 10:93–107.

Lachin JM; Matts JP; Wei LJ. (1988). Randomization in clinical trials: Conclusions and recommendations. *Control. Clin. Trials* 9:365–374.

Mehta CR; Patel NR; Senchaudhuri P; Tsiatis AA. (1994). Exact permutational tests for group sequential clinical trials. *Biometrics* 50:1042–1053.

Pesarin F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*. New York: Wiley.

Siegmund H. (1985). *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer.

Slud E; Wei LJ. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *JASA* 77:862–868.

Wei LJ; Smythe RT; Lin DY; Park TS. (1990). Statistical inference with data-dependent treatment allocation rules. *JASA* 85:156–162.

Chapter 15
Data Analysis

IN THIS CHAPTER WE REVIEW THE TOPICS you'll need to cover in your analysis of the data and the differing types of data you will encounter. For each type, you learn the best way to display and communicate results. You'll learn what analyses are performed, what tables and figures are generated, and what statistical procedures are employed for the analysis.

You'll walk step by step through the preparation of a typical final report. And you'll learn how to detect and avoid common errors in analysis and interpretation. A glossary of statistical terms is provided for help in decoding your statistician's reports.

REPORT COVERAGE

In this section we consider what material should be displayed and analyzed.

Each of the reports you prepare, from a brief abstract to the final comprehensive report, should cover the following topics:

- **Study population**
- **Baseline values**
- **Intermediate snapshots**
- **Protocol deviations**
- **Final results**
  - **— Primary end points**
  - **— Adverse events**
  - **— Other secondary end points**

The final comprehensive report also will have to include the following:

1. **Demonstrations of similarities and differences for**
    a. **Baseline values of the various treatment groups**
    b. **Data from the various treatment sites**
    c. **End points of the various subgroups determined by baseline variables and adjunct therapies**
2. **Explanations of protocol deviations including**
    a. **Ineligible patients who were accidentally included in the study**
    b. **Missing data**
    c. **Dropouts and withdrawals**
    d. **Modifications to treatment**

Further explanations and stratifications will be necessary if the frequencies of any of the protocol deviations differ among treatments. For example, if there are differences in the baseline demographics of the treatment groups, then subsequent results will need to be stratified accordingly. If one or two sites stand out from the rest, then the results from these must be analyzed separately. Moreover some plausible explanation for the differences must be advanced.

Here is another example. Suppose that the vast majority of women in the study were in the control group. Then, to avoid drawing false conclusions about the men, the results for men and women must be presented separately, unless you first can demonstrate that the treatments have similar effects on men and women.

UNDERSTANDING DATA

The way in which we present the data to be used in our reports and the methods of analysis we employ depend on the type of data involved.

As noted in Chapter 6, data may be divided into three categories:

1. **Categorical data such as sex and race.**
2. **Metric observations such as age where differences and ratios are meaningful.**
3. **Ordinal data such as subjective ratings of improvement, which may be viewed either as ordered categories or as discrete metric data depending on the context.**

In this preliminary section, we consider how we would go about displaying and analyzing each of these data types.

Categories

When we only have two categories as is the case with sex, we would report the number in one of the categories, the total number of

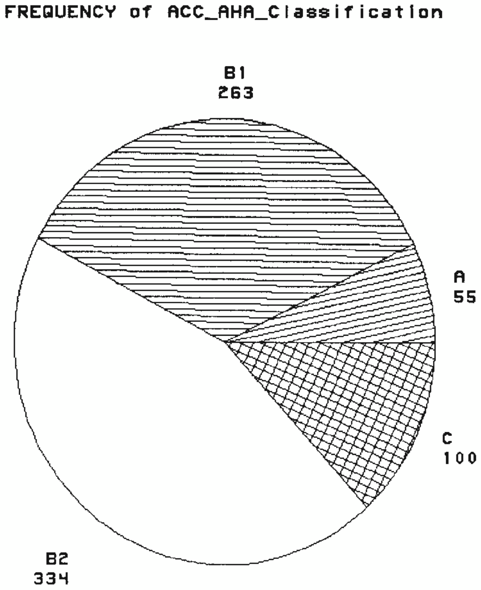


FIGURE 15.1 Pie Chart: Relative Proportions of Patients in Various ACA/AHA Classifications. The actual frequencies are also displayed.

meaningful observations, and the percentage as in "170 of 850 patients or 20% were females."

When we have four or more categories, the results are best summarized in the form of a pie chart as in Figure 15.1. Or, if we wish to make comparisons among multiple treatment groups, we can do this in the form of a banded bar chart as in Figure 15.2.

If there are only two treatments, we might also want to report a confidence interval for the *odds ratio* defined as $p_2(1 - p_1)/p_1(1 - p_2)$, where p_1 is the probability a subject in the first treatment group will belong to the first category and p_2 is a similar probability for the second treatment group.

If $p_2 = p_1$, the odds ratio is 1. If $p_2 > p_1$, the odds ratio is greater than 1. If the *confidence interval* for

CONFIDENCE INTERVALS

Bosses always want a bottom-line number. But the probability is zero that any single value based on clinical data will be correct. When caught between a boss and a hard place, the solution is to provide an interval, such as (0.98, 1.02), that you can guarantee with confidence will cover the correct value 90% or 95% of the time. As you might expect, the more precise and the greater the number of observations you make, the narrower that confidence interval will be.

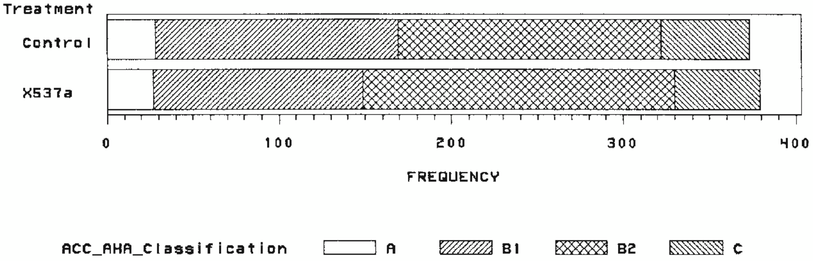


FIGURE 15.2 Bar Chart: Relative Proportions of Patients in Various ACA/AHA Classifications. The actual frequencies are also displayed.

the odds ratio includes 1, such as (0.98, 1.02), we have no reason to believe the two treatments have different effects on the variable we are studying.

Metric Data

For metric data such as age, we would normally report both the arithmetic mean of the sample and the standard error of the mean, for example, 59.3 ± 0.55 years, along with the sample size, $n = 350$. If the data take the form of a time to an event, it is more common to report the median or halfway point and to display the entire distribution in graphic form.

Appropriate Degree of Precision. Many computer programs yield values with eight or nine decimal places, most of which are meaningless. For example, because we can only measure age to the nearest day, it would be foolish to report mean age as 59.3724 years.

Even though we can measure age to the nearest day, it also would be foolish to report the mean age as 59.31 years, when the *standard error* is 0.55 The standard error is a measure of the *precision* of our estimate. It tells us how close we are likely to come to our original estimate if we were to repeat the sampling process many times.

If the underlying population has the form of a bell-shaped curve depicted in Figure 6.1, then in 95% of the samples we would expect the sample mean to lie within two standard errors of the mean of our original sample.

Increasing the sample size decreases the standard error, reflecting the increase in precision of the mean. By taking four times as many observations, we can cut the standard error in half. Had we made tens of thousands of observations in our hypothetical example, we would have been able to report the mean value as 59.31 ± 0.055.

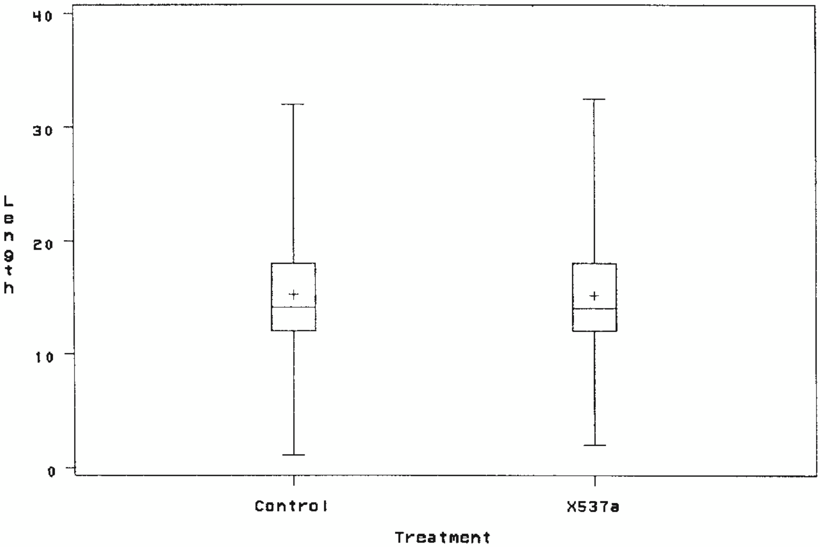


FIGURE 15.3 Box and Whiskers Plot. The box encompasses the middle 50% of each sample, while the "whiskers" lead to the smallest and largest values. The line through the box is the median of the sample; that is, 50% of the sample is larger than this value while 50% is smaller. The + indicates the sample mean. Note that the mean is shifted in the direction of a small number of very large values.

The standard error is not a measure of *accuracy*. I remember a cartoon depicting Robin Hood, bow in hand, examining where his arrows had each split the arrow in front of it. Unfortunately, all three arrows had hit a cow rather than the deer he was aiming at. The mean may not provide a valid representation of the center of the population when the observations do not come from a symmetric distribution such as that depicted in Figure 6.1.

When the data do not come from a symmetric distribution, it is preferable to report the median or 50th percentile along with the range and the 25th and 75th percentiles. Since it's hard to grasp so much information in text form, a box and whiskers plot such as that in Figure 15.3 provides the most effective way to present the data and to make a comparison between the two treatment groups.

If there are only two treatments, we might also want to report a confidence interval for the difference in mean values. If this confidence interval includes zero, we would infer that the treatments have approximately the same effect on the variable we are studying.

CHOOSING THE RIGHT STATISTIC

At first glance it would seem that statistics as a branch of mathematics ought to be an exact science and the choice of the correct statistical procedure determined automatically. But at least four influences are at work:

1. Accuracy. The *p*-value determined by a statistical method is correct (that is exact or accurate) only if the assumptions underlying the method are satisfied.
2. Computational feasibility. Rapid advances in hardware and software technology have made all but the most intractable of statistical methods practical today (e.g., see Good, 2000).
3. Regulatory agency requirements. The members of the various committees who exercise oversight on behalf of the regulatory agency must be satisfied with the statistical methods that are used. While counterarguments often fall on deaf ears, committee "recommendations" can often be forestalled by providing appropriate justification for the statistical techniques that are utilized, particularly when such techniques are a relatively recent introduction in the analysis of clinical trials.
4. Familiarity. Too often the choice of statistical method is determined on the basis of the technique that was used in the last set of clinical trials or the limited subset of techniques with which the biostatistician is familiar.

The fact that a method was not rejected in the past is no guarantee that it will not be rejected in the future. Regulatory agencies are composed of individuals. What one individual or individuals once found acceptable may meet with rejection by their replacements.

The only safety lies with carefully chosen proven statistical methodology.

Ordinal Data. When we have a small number of ordered categories (12 or less), the data should be reported in tabular form. Otherwise, report as you would metric data.

STATISTICAL ANALYSIS

How we conduct the analysis of the final results will depend on whether or not:

- **Baseline results of the treatment groups are equivalent**
- **Results of the disparate treatment sites may be combined**
- **(if adjunct treatments were used) Results of the various adjunct treatment groups may be combined**
- **Proportions of missing data, dropouts, and withdrawals are unrelated to treatment**

Thus the first steps in our analysis must be to address these issues.

TABLE 15.1 Sandoz Results by Test Site

	New Drug		Control Drug	
Test Site	Responded	*N*	Responded	*N*
2	0	39	6	32
3	1	20	3	18
4	1	14	2	15
5	1	20	2	19
6	0	12	2	10
7	3	49	10	2
8	0	19	2	17
9	1	14	0	15
10	2	26	2	27
11	0	19	2	18
12	0	12	1	11
13	0	24	5	19
14	2	10	2	11
15	0	14	11	3
16	0	53	4	48
19	1	50	1	48
20	0	13	1	13
21	0	13	1	13

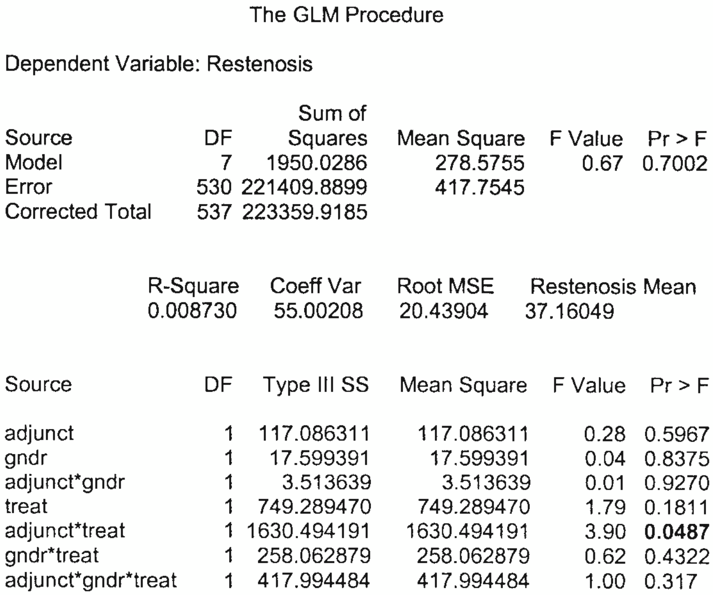
Consider the results summarized in Table 15.1 obtained by the Sandoz drug company and reproduced with permission from the StatXact manual. Obviously the results of sites 20 and 21 are the same, but are they the same as at the other sites? And what of site 15 with its extraordinarily large number of responses in the control group? Can the results for site 15 be combined with the results for the other sites? As it turns out, an analysis of these data with StatXact shows that there are statistically significant differences.

The analysis of a metric end point to determine if the data from various subsets may be combined usually takes the form of a *t-test* or an *analysis of variance* as in the sample output in Figure 15.4.

If we can resolve all the issues above in the negative, then the analysis is straightforward. Otherwise, we need to subdivide the sample into strata based on the differentiating factors and perform a separate analysis for each stratum.

Stratification may sometimes be necessitated even when the differences occasioned by differences in treatment are not statistically significant. See the section headed "Simpson's Paradox" later on in this chapter.

A recent clinical study illustrates some of the complications that can arise. Significant differences were found in the proportions of



```
                          The GLM Procedure

Dependent Variable: Restenosis

                                 Sum of
Source                DF        Squares    Mean Square   F Value   Pr > F
Model                  7      1950.0286       278.5755      0.67   0.7002
Error                530    221409.8899       417.7545
Corrected Total      537    223359.9185

             R-Square     Coeff Var     Root MSE     Restenosis Mean
             0.008730      55.00208     20.43904            37.16049


Source                DF    Type III SS    Mean Square   F Value   Pr > F

adjunct                1     117.086311     117.086311      0.28   0.5967
gndr                   1      17.599391      17.599391      0.04   0.8375
adjunct*gndr           1       3.513639       3.513639      0.01   0.9270
treat                  1     749.289470     749.289470      1.79   0.1811
adjunct*treat          1    1630.494191    1630.494191      3.90   0.0487
gndr*treat             1     258.062879     258.062879      0.62   0.4322
adjunct*gndr*treat     1     417.994484     417.994484      1.00   0.317
```

FIGURE 15.4 Results of an SAS Analysis. The analysis shows the joint effects of gender, adjunct therapy, and treatment on restenosis. The *p*-value (Pr > F) is less than 0.05 for just one of the terms above,[39] suggesting that the effect of adjunctive therapy on restenosis varies between the two treatment groups. A further detailed breakdown of the results reveals that while the adjunct therapy had a positive effect in the control group, its use was contradicted in the presence of the experimental treatment.

men and women that had been assigned to the various treatment groups. Exacerbating the situation was the discovery that men and women reacted differently to the adjuvant treatment.

The final results were broken out separately by men and women and whether they'd received the adjuvant or not (Table 15.2). One hundred percent of the women in the control group who received the adjuvant recovered completely, a totally unexpected result!

The adjuvant treatment also was of positive value for the men in the control group, but appeared to inhibit healing and was of negative value for those men who received the experimental treatment.

Categorical Data

Comparisons of categorical data may be displayed in the form of a contingency table. (Tables 15.1 and 15.3) In a 2 × 2 table such as that

---

[39] As we discuss further in what follows, these probabilities are at best approximations to the actual *p*-values.

TABLE 15.2 Binary Stenosis by Adjunct, Gender, and Treatment

		NuStent		Standard		
Adjunct	Gender	*N*	Mean	*N*	Mean	*p*
No	M	143	24	149	27	0.68
	F	51	26	42	24	0.97
Yes	M	67	28	60	17	0.15
	F	15	47	19	0	0.001

TABLE 15.3 Subset Analysis

	Full Recovery	Impairment
With adjunct	11	0
Without adjunct	17	5

of Table 15.3 the recommended analysis is Fisher's exact test. For a comparison of the odds ratios at various treatment sites as in Table 15.1, the recommended test is based on the permutation distribution of the Zelen statistic.

The chi-square distribution was used in the past to determine the significance level of both tables though it was well known that the chi-square distribution was only a poor approximation to the actual distribution of these statistics. For example, an analysis of Table 15.3 yields a *p*-value of 4.3% based on the chi-square distribution and Pearson's chi-square statistic, whereas the correct and exact *p*-value as determined by Fisher's method is 11.1%. An analysis of Table 15.1 yields a *p*-value of 7.8% based on the chi-square distribution, yet the correct and exact *p*-value as determined from the permutation distribution of Zelen's statistic is a highly significant 1.2%.

Today methods that yield exact and correct *p*-values for virtually every form of contingency table are available. See Mehta and Patel (1980, 1983, 1998), Mehta, Patel, and Tsiatis (1984), and Good (2001; Chapter 7). Yet many statisticians continue to utilize the erroneous chi-square approximation, much like a drunk might search for his missing wallet under the lamppost because the light there is better.

Ordinal Data

We often deal with data that are ordered but nonmetric such as self-evaluation scales. The observations can be ordered since "much improved" is obviously superior to merely "improved," but it is nonmetric because we cannot add and subtract the observations. To see that this is true, ask yourself whether one patient who is "much improved" and a second patient who shows "no change" are equivalent to two patients who are merely "improved?"

TABLE 15.4 Adverse Events per Patient by Treatment

	0	1	2	3	4	5	6	7	8	9	10	11	13
New	233	56	30	21	16	5	7	6	2	2	1	1	0
Standard	253	46	28	17	13	4	2	3	0	1	1	0	1

Such data have often been analyzed by chi-square methods as well. But a chi-square analysis is really a multi-sided test designed to detect any and all alternatives to the null hypothesis rather than being focused against the single-ordered alternative of interest. The result is that the chi-square analysis is not very powerful. If we analyze the data of Table 15.4 using the chi-square statistic, we will obtain a *p*-value of 0.51 and conclude that there is no significant difference between the two treatments. But if we note that the columns in the table are ordered from patients with no adverse events to patients with as many as 13 and use the Wilcoxon rank test, we can obtain a highly significant *p*-value of 0.025.[40]

Metric Data

Statistics is an evolving science. Statisticians are always trying to develop new and more powerful statistical techniques that make the most from the data at hand.

Virtually all statistical tests employed in clinical trials require that:

1. **Patients be drawn at random from the population**
2. **Patients be assigned at random to treatment**
3. **Observations on different patients be independent of one another**
4. **Under the hypothesis of no treatment differences, *the null hypothesis*, all the observations in the samples being compared come from the same distribution**[41]

Parametric methods also require that the observations come from a normal distribution, that is, a bell-shaped curve similar to the one depicted in Figure 6.1. Thus nonparametric tests that do not have this restriction are usually to be preferred to parametric for the analysis of metric data.

Examples of *parametric* methods include the *t*-test for comparing means and the *F*-test used in the analysis of variance. The *F*-test provides exact p-values if all the above restrictions are met. But even with only moderate deviations from the bell-shaped normal distribution, the p-values provided by the F-test can be quite misleading. The t-test is more robust and provides almost exact p-values for most samples of metric data larger than 10 in number.[42]

---

[40] Even this latter test may be improved upon; see Berger, Permutt, and Ivanova (1998).

[41] Implicit to this assumption is that the patients have been randomly assigned to treatment.

Examples of *nonparametric* methods include permutation tests based on rank, such as the Wilcoxon test for comparing two samples and the Kruskall-Wallace test for comparing k samples, and permutation tests based on the original observations. These tests always provide exact significance levels if the two basic assumptions of independence and equal variances are satisfied.

Permutation tests based on the original observations are more powerful and should be used in preference to rank tests unless there is reason to believe the data may contain one or more *outliers* (exceptional values or typographical errors). Ranks diminish the effects of outliers. For example, the mean of 1.1, 2.2, 3.4, 4.3, 59, is 14; in taking ranks, the mean of 1, 2, 3, 4, 5, is 3.

Software providing for permutation tests based on the original observations has been in short supply until recently. Today permutation software is available for comparing two samples, for comparing k samples with either ordered or unordered categories, and for the analysis of multifactor designs.

A second weakness of the parametric analysis of variance approach is that it only provides for a test of the null or no-difference-in-effect hypothesis against all possible alternatives. If we know in advance that the alternative has a specific form (e.g., an ordered dose response), then we can always find a more powerful permutation test to take advantage of this knowledge. For more on this topic, see Salsburg (1992).

An Example. Owning a statistics program will no more makes you a statistician, then buying a pamphlet called "Brain Surgery Made Easy" will turn you into a neurosurgeon. While almost anyone can learn to use a statistics program (or a scalpel), interpreting the results is quite a different matter. Consider the output of one of the less complex of SASs many statistics routines, the Ttest procedure:

---

[42] Of course, one easily can find extreme cases that are an exception to this rule.

```
                         The TTEST Procedure

                              Statistics

                         Lower   Upper
                         CL      CL         Lower CL         Upper CL
Variable treat N    Mean    Mean    Mean    Std Dev Std Dev Std Dev Std Err
RIG      New 121     0.5527 0.5993 0.6459 0.2299  0.2589  0.2964  0.0235
RIG      Stand 127   0.5721 0.598  0.6238 0.1312  0.1474  0.1681  0.0131
RIG      Diff (1–2) –0.051  0.0013 0.0537 0.1924  0.2093  0.2296  0.0266

                               T-Tests

Variable     Method            Variances     DF      t Value      Pr > |t|
RIG          Pooled            Equal         246     0.05         0.9608
RIG          Satterthwaite     Unequal       188     0.05         0.9613

                        Equality of Variances

Variable      Method         Num DF       Den DF       F Value       Pr > F
RIG           Folded F       120          126          3.09          <.0001
```

The first table of the SAS output provides us with confidence limits for the mean and standard deviation of the variable RIG for each of the two treatment groups. We would report the results as RIG is 0.59 ± 0.02 for those receiving the new treatment and 0.59 ± 0.01 for those receiving the standard.

There does not appear to be a significant difference between the RIG values of the two treatments for the *t*-value is quite small and the probability of observing a larger *t*-value by chance is 0.96 or close to 1. However, because the variances are significantly different (as shown in the last row of the output) the results of the *t*-test shown in the second table cannot be relied on.[43]

Time-to-Event Data

Possible events you might wish to track the time after treatment to include "symptom free," "recurrence of symptoms," and "death." While survival or time-to-event data are metric (or at least ordinal), the presentation and analysis of results takes on a quite different character. If the events are inevitable and the trials last long enough, then we can compare treatments as we do with metric data using either a permutation test applied to the original observations or, if we want to diminish the effects of a few very lengthy time intervals, the ranks of the observations in the combined sample.[44]

---

[43] Details of the correct statistical procedure to use with unequal variances, known as the Behrens-Fisher problem, are given later in this chapter.

[44] Tests based on a normal distribution would not be appropriate as the distribution is far from normal, the mean "time-to-event" typically being much greater than the median.

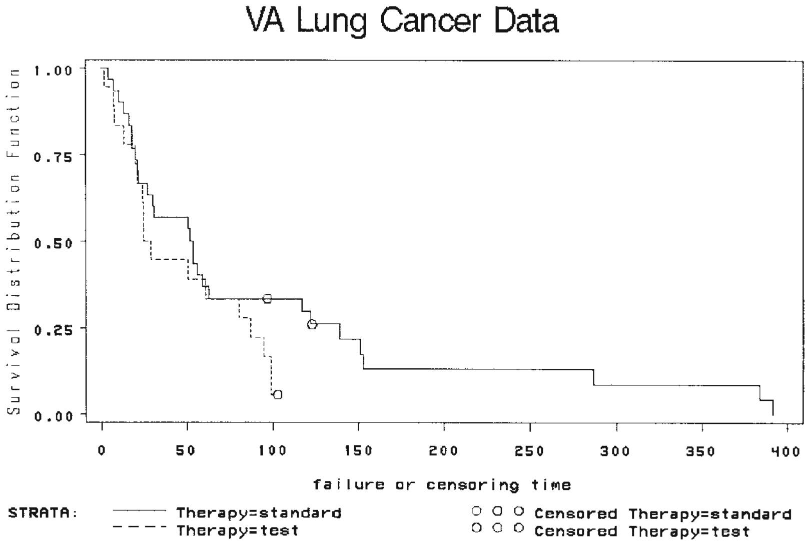


FIGURE 15.5 Depicting Time-to-Event Data with the Aid of a Survival Curve.

But time-to-event data are often censored; for many of the patients, the event being tracked may have not yet occurred by the time the trials end, so only a minimum value can be recorded. A graph such as Figure 15.5 is the most effective way to present the results. The circles denote those observations that are censored; they represent times that might have been much longer had the trials been allowed to continue.

In most animal experiments, all the subjects receive the treatment on the same date and are subject to the same degree of censoring. The optimal statistic, described by Good (1992), takes into account both the time to event for those animals in which the event occurred during the study period, and the relative proportions between treatments of animals who complete the trials without the event occurring.

In trials with humans, patients are enrolled on an ongoing basis. One patient may be followed for several years, while another may be enrolled in the trials only a few months before they end. Patients who enter the study long after the trials have begun are more likely to have small recorded values. This is one of the reasons why we often specify two cut-off dates for trials, the earlier denoting the last date on which a patient may enter the study and the later representing the date on which the last observation will be made.

TABLE 15.5 SAS Output from a Proc Lifetest Analysis

```
               Univariate Chi-Squares for the Wilcoxon Test
            Test         Standard       Pr>            
Variable    Statistic    Deviation      Chi-Square     Chi-Square
Treatment   −1.9670      1.9399         1.0281         0.3106
               Univariate Chi-Squares for the Log-Rank Test
            Test         Standard       Pr>
Variable    Statistic    Deviation      Chi-Square     Chi-Square
Treatment   −4.3108      2.8799         2.2405         0.1344
          Forward Stepwise Sequence of Chi-Squares for the Wilcoxon Test
                         Pr>            Chi-Square     Pr>
Variable    DF           Chi-Square     Chi-Square     Increment      Increment
Karnofsky   1            11.0918        0.0009         11.0918        0.0009
  index
Treatment   2            11.4047        0.0033         0.3128         0.5759
          Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test
                         Pr>            Chi-Square     Pr>
Variable    DF           Chi-Square     Chi-Square     Increment      Increment
Karnofsky   1            5.4953         0.0191         5.4953         0.0191
  index
Treatment   2            7.9177         0.0191         2.4224         0.1196
```

A different form of analysis is called for, one that imputes values to the censored observations based on a mathematical model. The two principal methods are a log-rank test and a test based on a censored Wilcoxon.

Generally, the log-rank test should be employed if the emphasis is to be placed on early time-to-event data. The results for the data depicted in Figure 15.5 are given in Table 15.5. The large *p*-value of 0.3 or 30% reveals that treatment does not have a significant effect on survival.

In many cases one would also want to correct for cofactors. The second part of Table 15.5 reveals the statistically significant relation of survival to the Karnofsky Index which is a measure of the overall status of the cancer patient at the time of entry into the clinical trials.

TYPES OF DATA

Binomial. Observations fall into one of two categories, heads versus tails, success versus failure, yes versus no.

Categorical. Data are subdivided into categories such as black, white, Hispanic.

Ordinal. Observations can be ordered from smallest to largest (though there may be ties). Examples include rating scales.

Metric. Ordinal observations are meaningful for their differences. Examples include age, height, and percent stenosis.

Survival. Time to the event is recorded. Examples include survival time, time to relapse, and time till the absence of a symptom or symptoms.

STEP BY STEP

The guidelines presented here are generic; always consult the current guidelines published by the cognizant regulatory agency before beginning or submitting the results of a statistical analysis. A copy of the 1996 IHS Guideline for the Format and Content of the Clinical and Statistical Sections of an Application may be downloaded from *http://www.fda.gov/cder/guidance/statnda.pdf*. Guidelines for individual European nations may be accessed via *http://www.eudra.org/en_home.htm*.

The purpose of the statistical analysis section of the final report like that of the final report itself is tell a story in as clear and concise a fashion as possible. Only the tabular material and graphs associated with the main theme should be incorporated directly in the analysis section. All the other material described in what follows should be deferred to appendexes.

The Study Population

Begin by describing the study population in terms of the eligibility criteria, the proportion of males and females, average age of the participants, and so forth. For categorical variables such as sex and race, report both the percentages and numbers in each category. For continuous variables such as age, report the mean, standard error of the mean, and sample size.

Note that sample and subsample sizes are always based on the number of patients, not on the number of eyes or tumors or teeth or injuries that may have been treated. This is true even if the individual organs have been treated independently of one another.

A side-by-side comparison of baseline values should be provided in tabular form as in Table 15.6. Include confidence limits, but do not include *p*-values. In the opinion of Lang and Secic (1997), "Clinical imbalances, even though the result of chance, are real and need to be accounted for in the subsequent analysis. Statistical comparisons of baseline values are rarely necessary.

TABLE 15.6 Baseline Comparison

	New, $N = 380$	Standard, $N = 370$	Odds Ratio
Men	74%	78%	0.80 (0.57, 1.12)
Diabetes	36%	35%	0.96 (0.70, 1.30)
Current smoker	28%	29%	1.04 (0.75, 1.44)
Elevated cholesterol	58%	58%	0.99 (0.73, 1.34)
Hypertension	59%	54%	0.80 (0.60, 1.09)

"Statistically significant differences will be the result of chance [because of the large number of simultaneous tests that are being performed], and nonstatistically significant differences do not indicate that the groups are comparable, but rather that randomization was effective." See also Altman and Dore (1990).

Show that the results of the various treatment sites can be combined. If the end point is binary in nature—success or failure—employ Zelen's (1971) test of equivalent odds ratios in 2 × 2 tables. If it appears that one or more treatment sites should be excluded, provide a detailed explanation for the exclusion if possible ("repeated protocol violations," "ineligible patients," "no control patients," "misdiagnosis") and exclude these sites from the subsequent analysis.[45]

Determine which baseline and environmental factors, if any, are correlated with the primary end point. Perform a statistical test to see if there is a differential effect between treatments as a result of these factors. Test to see if there is a differential effect on the end point among treatments occasioned by the use of any adjunct treatments.

Reporting Primary End Points

Report the results for each primary endpoint separately. For each endpoint:

1. **Report the aggregate results by treatment for all patients who were examined during the study.**
2. **Report the aggregate results by treatment only for those patients who were actually eligible, who were treated originally as randomized, and who were not excluded for any other reason. Provide *p*-values for treatment comparisons.**
3. **Break down these latter results into subsets based on factors determined prior to the start of the study such as adjunct therapy or gender. Provide *p*-values for treatment comparisons.**
4. **List all factors uncovered during the trials that appear to have altered the effects of treatment. Provide a tabular comparison by treatment for these factors but do not include *p*-values.**

If there were multiple endpoints, you have the option of providing a further multivariate comparison of the treatments.

---

[45] Any explanation is bound to trigger inquiries from the regulatory agency. This is yet another reason why continuous monitoring of results and subsequent early remedial action are essential.

Exceptions

Every set of large-scale clinical trials has its exceptions. You must report the raw numbers of such exceptions and, in some instances, provide additional analyses that analyze or compensate for them. Typical exceptions include the following:

Did Not Participate. Subjects who were eligible and available but did not participate in the study—this group should be broken down further into those who were approached but chose not to participate and those who were not approached.

Ineligibles. In some instances, depending on the condition being treated, it may be necessary to begin treatment prior to ascertaining whether the subject was eligible to participate in the study.

For example, an individual arrives at a study center in critical condition. The study protocol calls for a series of tests, whose results may not be back for several days, but in the opinion of the examining physician, treatment must begin immediately. The patient is randomized to treatment and only later is it determined that the patient is ineligible.

The solution is to present two forms of the final analysis, one incorporating all patients, the other limited to those who were actually eligible.

Withdrawals. Subjects who enrolled in the study but did not complete it. Includes dropouts and noncompliant patients. These patients might be subdivided further based on the point in the study at which they dropped out.

At issue is whether such withdrawals were treatment related. For example, the gastro-intestinal side effects associated with erythromycin are such that many patients (including me) may refuse to continue with the drug.

If possible, subsets of both groups should be given detailed follow-up examinations to determine whether the reason for the withdrawal was treatment related.

Crossovers. If the design provides for intent-to-treat, a noncompliant patient may still continue in the study after being reassigned to an alternate treatment. Two sets of results should be reported: one for all patients who completed the trials (retaining their original assignments), and one only for those patients who persisted in the groups to which they were originally assigned.

Missing Data. Missing data is a common, expensive, and preventable problem in many instances.

The primary end point of a recent clinical study of various cardiovascular techniques was based on the analysis of follow-up angiograms. Although more than 750 patients had been enrolled in the study, only 523 had the necessary angiograms. Put another way, almost a third of the monies spent on the trials had been wasted.

Missing data are often the result of missed follow-up appointments. The recovering patient no longer feels the need to return or, at the other extreme, is too sick to come into the physician's office. Noncompliant patients are also likely to skip visits.

You need to analyze the data to ensure that the proportions of missing observations are the same in all treatment groups. If the observations are critical, involving primary or secondary end points as in the preceding example, then you will need to organize a follow-up survey of at least some of the patients with missing data. Such surveys are extremely expensive.

As always, prevention is the best and sometimes the only way to limit the impact of missing data.

- **Ongoing monitoring and tying payment to delivery of critical documents are essential.**
- **Site coordinators on your payroll rather than the investigator's are more likely to do immediate follow-up when a patient does not appear at the scheduled time.**
- **A partial recoupment of the missing data can be made by conducting a secondary analysis based on the most recent follow-up value. See Pledger (1992).**

A chart such as that depicted in Figure 15.6 is often the most effective way to communicate all this information; see, for example, Lang and Secic (1997, p. 22).

Outliers. Suspect data may be reported such as depicted in Figure 14.2. You may want to perform two analyses, one incorporating all the data, and one deleting the suspect data. A further issue is whether the proportion of suspect data is the same for all treatment groups.

Competing Events. A death or a disabling accident, whether or not it is directly related to the condition being treated, may prevent us from obtaining the information we need. The problem is a common one in long-term trials in the elderly or high-risk populations and is best compensated for by taking a larger sample to begin with.

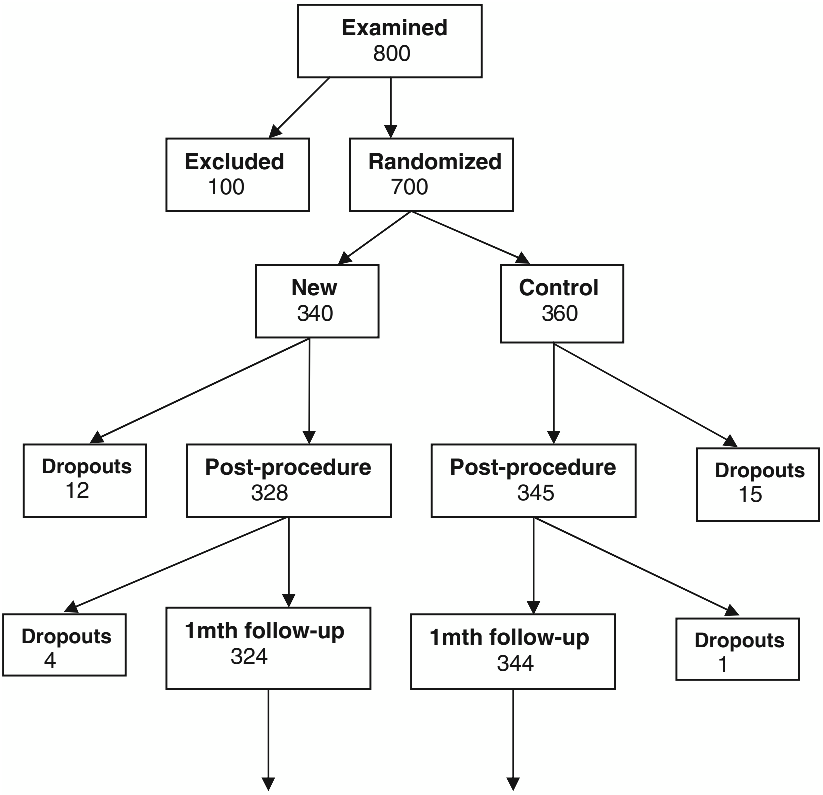

Examined
800
Excluded
100
Randomized
700
New
340
Control
360
Dropouts
12
Post-procedure
328
Post-procedure
345
Dropouts
15
Dropouts
4
1mth follow-up
324
1mth follow-up
344
Dropouts
1

FIGURE 15.6 Where Did All the Patients Go?

Adverse Events

Report the number, percentage, and type of adverse events associated with each treatment. Accompany this tabulation with a statistical analysis of the set of adverse events as a whole as well as supplementary analyses of classes of adverse events that are known from past studies to be treatment or disease specific.

Report the incidence of adverse events over time as a function of treatment. Detail both changes in the total number of adverse events and in the number of patients who remain incident free. You may also wish to distinguish various levels of severity.

ANALYTICAL ALTERNATIVES

In this section we consider some of the more technically challenging statistical issues on which statisticians often cannot agree including

(1) unequal variances, (2) testing for equivalence, (3) Simpson's paradox, and (4) estimating precision.

When Statisticians Can't Agree

Statistics is not an exact science. Nothing demonstrates this more than the Behrens-Fisher problem of unequal variances in the treatment groups. Recall that the *t*-test for comparing results in two treatment groups is valid only if the variances in the two groups are equal. Statisticians do not agree on which statistical procedure should be used if they are not. When I submitted this issue recently to a group of experienced statisticians, almost everyone had their own preferred method. Here is just a sampling of the choices:

- ***t*-test. One statistician commented, "SAS PROC TTEST is nice enough to present *p*-values for both equal and unequal variances. My experience is that the FDA will always accept results of the *t*-test without the equal variances assumption—they would rather do this than think."**
- **Wilcoxon test. The use of the ranks in the combined sample reduces the impact (though it does not eliminate the effect) of the difference in variability between the two samples.**
- **Generalized Wilcoxon test, see O'Brien (1988).**
- **Procedure described in Manly and Francis (1999).**
- **Procedure described in Chapter 7 of Weerahandi (1995).**
- **Procedure described in Chapter 10 of Pesarin (2001).**
- **Bootstrap. Draw two bootstrap samples independently from each sample; compute the mean and variance of each bootstrap sample. Compute the *t*-statistic. Repeat to derive a bootstrap distribution for the *t*-statistic. Reject if the original value of the test statistic is an extreme value of this distribution.**

Hilton (1996) compared the powers of the Wilcoxon test, O'Brien's test, and the Smirnov test in the presence of both location shift and scale (variance) alternatives. As the relative influence of the difference in variances grows, the O'Brien test is most powerful. The Wilcoxon test loses power in the face of different variances. If the variance ratio is 4:1, the Wilcoxon test is virtually useless.

One point is unequivocal. In a personal communication, William Anderson advises', "The first issue is to understand *why* the variances are so different, and what does this mean to the patient. It may well be the case that a new treatment is not appropriate because of higher variance, even if the difference in means is favorable. This issue is important whether or not the difference was anticipated. Even if the regulatory agency does not raise the issue, I want to do so internally."

David Salsburg agrees. "If patients have been assigned at random to the various treatment groups, the existence of a significant difference in any parameter of the distribution suggests that there is a difference in treatment effect. The problem is not how to compare the means but how to determine what aspect of this difference is relevant to the purpose of the study.

> **"Since the variances are significantly different, I can think of two situations where this might occur:**
>
> 1. **In many clinical measurements there are minimum and maximum values that are possible, e.g., the Hamilton Depression Scale, or the number of painful joints in arthritis. If one of the treatments is very effective, it will tend to push patient values into one of the extremes. This will produce a change in distribution from a relatively symmetric one to a skewed one, with a corresponding change in variance.**
> 2. **The patients may represent a mixture of populations. The difference in variance may occur because the effective treatment is effective for only a subset of the patient population. A locally most powerful test is given in Conover and Salsburg (1988)."**

Testing for Equivalence

The statistical procedures for testing for statistical significance and for equivalence are quite different in nature.

The difference between the observations arising from two treatments T and C is judged ***statistically significant*** if it can be said with confidence level α that the difference between the mean effects of the two treatments is greater than zero.

Another way of demonstrating precisely the same thing is to show $c_L \le 0 \le c_R$ where c_L and c_R are the left and right boundaries, respectively, of a $1 - 2\alpha$ confidence interval for the difference in treatment means.

The value of α is taken most often to be 5% ($\alpha = 10\%$ is sometimes used in preliminary studies). In some instances, such as ruling out adverse effects, 1% or 2% may be required.

Failure to conclude significance does not mean that the variables are equal, or even equivalent. It may merely be the result of a small sample size. If the sample size is large enough, any two variables will be judged significantly different.

Two treatments T and C will be called *equivalent* if the difference between the mean effects of the two treatments is less than a value Δ, called the *minimum relevant difference.*

TABLE 15.7 Equivalence versus Statistical Significance

	−	Δ 0	+	Δ
Equivalent Not statistically significant		(	)	
Equivalent Statistically significant			( )	
Not equivalent Not statistically significant		(		)
Not equivalent Statistically significant			(	)

This value Δ is chosen based on clinical, engineering, or scientific reasoning. There is no traditional mathematical value.

To perform a test of equivalence, we need to generate a confidence interval for the difference of the means:

1. **Choose a sample from each group.**
2. **Construct a confidence interval for the difference of the means. For significance level α, this will be a $1 - 2\alpha$ confidence interval.**
3. **If $-\Delta \le c_L$ and $c_R \le \Delta$, the groups are judged equivalent.**

Table 15.7 depicts the left "("and right")" boundaries of such a confidence interval in a variety of situations.

Failure to detect a significance difference does not mean that the treatment effects are equal, or even equivalent. It may merely be the result of a small sample size. If the sample size is large enough, any two samples will be judged significantly different.

Simpson's Paradox

A significant *p*-value in the analysis of contingency tables only means that the variables are associated. It does not mean there is a cause and effect relationship between them. They may both depend on a third variable omitted from the study.

Regrettably, a third omitted variable may also result in two variables appearing to be independent when the opposite is true. Consider the following table, an example of what is termed Simpson's paradox:

Population

	Control	*Treated*
Alive	**6**	**20**
Dead	**6**	**20**

We don’t need a computer program to tell us the treatment has no effect on the death rate. Or does it? Consider the following two tables that result when we examine the males and females separately:

Males

	Control	*Treated*
Alive	**4**	**8**
Dead	**3**	**5**

Females

	Control	*Treated*
Alive	**2**	**12**
Dead	**3**	**15**

In the first of these tables, treatment reduces the male death rate from 0.43 to 0.38. In the second from 0.6 to 0.55. Both sexes show a reduction, yet the combined population does not. Resolution of this paradox is accomplished by avoiding a knee jerk response to statistical significance when association is involved. One needs to think deeply about underlying cause and effect relationships before analyzing data. Thinking about cause and effect relationships in the preceding example might have led us to thinking about possible sexual differences, and to testing for a common odds ratio.

Estimating Precision

Reporting results in terms of a mean and standard error as in 56 ± 3.2 is a long-standing tradition. Indeed, many members of regulatory committees would protest were you to do otherwise. Still, mathematical rigor and not tradition ought prevail when statistics is applied. Rigorous methods for estimating the precision of a statistic include transformation/back-transformation and the bootstrap.

When metric observations come from a bell-shaped symmetric distribution, the probability is 95% on the average that the mean of the population lies within two standard errors of the sample mean. But if the distribution is not symmetric, as is the case when measurement errors are a percentage of the measurement, then a nonsymmetric interval is called for. One first takes the logarithms of the observations, computes the mean and standard error of the logarithms, and determines a symmetric confidence interval. One then takes the antilogarithms of the boundaries of the confidence interval and uses these to obtain a confidence interval for the means of the original observations.

The drawback of the preceding method is that it relies on the assumption that the distribution of the logarithms is a bell-shaped distribution. If it is not, we’re back to square one.

With the large samples that characterize long-term trials, the use of the *bootstrap* is always preferable. When we bootstrap, we treat the original sample as a stand-in for the population and resample from it repeatedly, 1000 times or so, with replacement, computing the average each time.

For example, here are the heights of a group of adolescents, measured in centimeters and ordered from shortest to tallest.

137.0 138.5 140.0 141.0 142.0 143.5 145.0 147.0 148.5 150.0 153.0 154.0
155.0 156.5 157.0 158.0 158.5 159.0 160.5 161.0 162.0 167.5

The median height lies somewhere between 153 and 154 centimeters. If we want to extend this result to the population, we need an estimate of the precision of this average.

Our first bootstrap sample, which I've arranged in increasing order of magnitude for ease in reading, might look like this:

138.5 138.5 140.0 141.0 141.0 143.5 145.0 147.0 148.5 150.0 153.0 154.0
155.0 156.5 157.0 158.5 159.0 159.0 159.0 160.5 161.0 162.0

Several of the values have been repeated as we are sampling *with replacement*. The minimum of this sample is 138.5, higher than that of the original sample; the maximum at 162.0 is less than the original, while the median remains unchanged at 153.5.

137.0 138.5 138.5 141.0 141.0 142.0 143.5 145.0 145.0 147.0 148.5 148.5 150.0
150.0 153.0 155.0 158.0 158.5 160.5 160.5 161.0 167.5

In this second bootstrap sample, we again find repeated values; this time the minimum, maximum, and median are 137.0, 167.5, and 148.5, respectively.

The medians of 50 bootstrapped samples drawn from our sample ranged between 142.25 and 158.25 with a median of 152.75 (see Figure 15.7). They provide a feel for what might have been had we sampled repeatedly from the original population.

The bootstrap may also be used for tests of hypotheses. See, for example, Freedman et al. (1989) and Good (2001, Chapter 5).

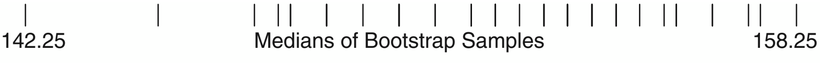


FIGURE 15.7 Scatterplot of 50 Bootstrap Medians Derived from a Sample of Heights.

BAD STATISTICS

Among the erroneous statistical procedures we consider in what follows are

- **Using the wrong method**
- **Choosing the most favorable statistic**
- **Making repeated tests on the same data (which also was considered in the previous chapter)**
- **Testing ad hoc, post hoc hypotheses.**

Using the Wrong Method

The use of the wrong statistical method—a large-sample approximation instead of an exact procedure, a multipurpose test instead of a more powerful one focused against specific alternatives, or a test whose underlying assumptions are clearly violated—can, in most instances, be attributed to what Peddiwell and Benjamin (1959) term the Saber-Tooth Curriculum. Most statisticians were taught already outmoded statistical procedures and too many haven't caught up since.

A major recommendation for your statisticians (besides making sure they have copies this and all my other books) is that they remain current with evolving statistical practice. Continuing education, attendance at meetings and conferences directed at statisticians, as well as seminars at local universities and think tanks are musts.

STATISTIC CHECKLIST

Is the method appropriate to the type of data being analyzed?

Should the data be rescaled, truncated, or transformed prior to the analysis?

Are the assumptions for the test satisfied?

- Samples randomly selected
- Observations independent of one another
- Under the no-difference or null hypothesis, all observations come from the same theoretical distribution
- (Parametric tests) The observations come from a specific distribution
- Is a more powerful test statistic available?

Choosing the Most Favorable Statistic

Earlier we saw that one might have a choice of several different statistics in any given testing situation. Your choice should be spelled out in the protocol. It is tempting to choose among statistics and data transformations after the fact, selecting the one that yields or comes closest to yielding the desired result. Such a "choose-the-best" procedure will alter the stated significance level and is unethical.

Making Repeated Tests on the Same Data

In the International Study of Infarct Survival (1988), patients born under the Gemini or Libra astrological birth signs did somewhat worse on aspirin than no aspirin in contrast to the apparent beneficial effects of aspirin on all other study participants.

Alas for those nutters of astrological bent, there is no hidden meaning in this result. When we describe a test as significant at the 5% or 1 in 20 level, we mean that one in 20 times, we'll get a significant result by chance alone. That is, when we test to see if there are any differences in the baseline values of the control and treatment groups, if we've made 20 different measurements, we can expect to see at least one statistically significant difference. This difference will not represent a flaw in our design but simply chance at work. To avoid this undesirable result—that is, to avoid making a Type I error and attributing to a random event an effect where none exists—we have three alternatives:

1. **Using a stricter criteria for statistical significance, 1 in 50 times (2%) or 1 in 100 (1%) instead of 1 in 20 (5%).**
2. **Applying a correction factor such as that of Bonferroni that automatically applies a stricter significance level based on the number of tests we've made.**
3. **Distinguishing among the hypotheses we began the study with (and accepting or rejecting these at the original significance level) while demanding additional corroborating evidence for those exceptional results (e.g., a dependence on astrological sign) that are uncovered for the first time during the trials.**

Which alternative you adopt will depend on the underlying situation.

If you have measured 20 or so study variables, then you will make 20 not entirely independent comparisons, and the Bonferroni inequality or the Westfall sequential permutation procedure is recommended.

If you are performing secondary analyses of relations observed after the data were collected, that is, relations not envisioned in the original design, then you have a right to be skeptical and to insist on either a higher significance level or to view the results as tentative requiring further corroboration.

A second example in which we have to modify rejection criteria is the case of adaptive testing that we considered in the previous chapter. To see why we cannot use the same values to determine statistical significance when we make multiple tests that we use for a single nonsequential test, consider a strategy many of us adopt when we play with our children.

You are playing the best out of three games. If your child wins, you call it a day. If you win, you say let's play three out of five. If you win the next series, then you make it four out of seven, and so forth. In most cases, by the time you quit, your child is able to say to his mother, "I beat daddy."[46]

Increasing the number of opportunities one has to win or to reject a hypothesis shifts the odds, so that to make the game fair again, or the significance level accurate, the rejection criteria must be modified.

Ad hoc, Post hoc Hypotheses

Patterns in data can suggest but cannot confirm hypotheses unless these hypotheses were formulated *before* the data were collected. Everywhere we look, there are patterns. In fact, the harder we look the more patterns we see. It is natural for us to want to attribute some underlying cause to these patterns. But those who have studied the laws of probability tell us that more often then not patterns are simply the result of random events. Put another way, a cluster of events in time or in space has a greater probability than equally spaced events.[47]

How can we determine whether an observed association represents an underlying cause and effect relationship or is merely the result of chance? The answer lies in the very controlled clinical trials we are conducting. When we set out to test a specific hypothesis, then the probability of a specific event is predetermined. But when we uncover an apparent association, one that may well have arisen purely by chance, we cannot be sure of the association's validity until we conduct a second set of controlled clinical trials.

In this section we look at three examples taken (with suitable modifications to conceal their identity) from actual clinical trials.

Random, Representative Samples. The purpose of a recent set of clinical trials was to see if a simple surgical procedure performed prior to taking a standard prescription medicine would improve blood flow and distribution in the lower leg.

The results were disappointing on the whole, but one of the marketing representatives noted that when a marked increase in blood flow was observed just after surgery, the long-term prognosis was excellent. She suggested that we calculate a p-value for a comparison of patients with an improved blood flow versus patients who had taken the prescription medicine alone.

---

[46] With teenagers, we sometimes try to make this strategy work in our favor.

[47] See, for example, Good (2001a, Chapter 13).

Such a *p*-value would be meaningless. Only one of the two samples of patients in question had been taken at random from the population. The other sample was determined after the fact. In order to extrapolate results from the samples in hand to a larger population, the samples must be taken at random from and be representative of that population.[48]

An examination of surgical procedures and of those characteristics that might forecast successful surgery definitely was called for. But the generation of a *p*-value and the drawing of any final conclusions must wait on clinical trials specifically designed for that purpose.

Finding Predictors. A logistic regression reveals the apparent importance of certain unexpected factors in a trial's outcome including gender. A further examination of the data reveals that the 16 female patients treated with the standard therapy and the adjunct all realized a 100% recovery. Because of the small numbers of patients involved, and the fact that the effects of gender were not one of the original hypotheses, we cannot report a *p*-value. But we should consider launching a further set of trials targeted specifically at female patients.

We need to report all results to the regulatory agency separately by sex as well as with both sexes combined. We need to research the literature to see if there are prior reports of dependence on sex.[49] Not least, we need to perform a cost–benefit analysis to see if a set of clinical trials using a larger number of female subjects would be warranted. (See also Chapter 16.)

Adverse Events. We'd been fortunate in that only a single patient had died during the first six months of trials, when we received a report of a second death. Although, over 30 sites were participating in the trials, the second death was at the same clinic as the first! The multiple deaths warranted an intensive investigation of procedures at that clinic, even though we could not exclude the possibility that chance alone was at work.

---

[48] See Section 2.7 of the second edition of Good (2001b) for a more detailed discussion.

[49] We found that the original clinical trials of the adjunct had used only men.

INTERPRETATION

The last example in the preceding section illustrates the gap between statistical and clinical significance.

Statistical significance is defined as the probability that an event might have occurred by chance alone. The smaller the probability, the more statistically significant the result is said to be. *Clinical significance* is that aspect of a study's outcome that convinces physicians to modify or maintain their current practice of medicine. Statistical significance can be a factor in determining clinical significance, but only if it occurs in conjunction with other clear and convincing evidence.[50]

Don't pad your reports and oral presentations with clinically insignificant findings. Do report statistically insignificant differences if the finding is of clinical significance.

Consider expressing the results of the primary outcome measures in the most clinically significant fashion. For example, on a practical day-to-day level, it is the individual who concerns us, not the population. I don't care about mean blood levels when I have a headache; I want to know whether *my* headache will get better. The percentage of patients who experienced relief or a total cure will be more meaningful to me than any average. For example, in reports on cardiovascular surgery, it is customary to report the rate of binary restenosis (occlusion of a coronary artery in excess of 50%) in the sample, along with the mean value for arterial occlusion.

SOFTWARE DOCUMENTATION

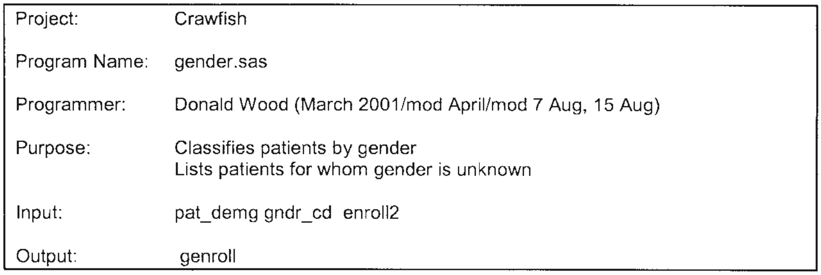
As noted in previous chapters, the programs that can be used for interim analyses can also be used for analysis of the final results. Thus development of the programs used for analysis should begin at the same time as or just prior to completion of the programs used for data entry.

Two sets of programs are needed, one for the extraction of data, and one for statistical analysis.

Insist on documentation of all computer programs used for data selection and analysis. Programmers as much or more as other staff in your employ tend to be highly mobile. You cannot rely on programmers who were with you during the early stages of the trial to be present at the trial's conclusion. Your staff should be encouraged

---

[50] See *http://www.indiana.edu/~stigtsts/* and MacKay and Oldford (2001) for more on this topic.



Project:	Crawfish
Program Name:	gender.sas
Programmer:	Donald Wood (March 2001/mod April/mod 7 Aug, 15 Aug)
Purpose:	Classifies patients by gender Lists patients for whom gender is unknown
Input:	pat_demg gndr_cd enroll2
Output:	genroll

FIGURE 15.8 The Header. The header briefly summarizes the essential information about each program including the name of the programmer, the program's purpose, the files it requires for input, and the files it creates.

TABLE 15.8 SAS Programs Used in the Analysis of Crawfish (does not include ad hoc queries)

	Developed by Donald Wood	13-Aug-01	
File Name	Purpose	Input	Output
adverse	Calculates frequency of adverse events	adv_evnt random	mace
age02	Calculates age of patients; prints list of patno's with misg data.	pat_demg, smlsrg, enroll2	
angio	Computes average preprocedure lesion length Prints list of patients without CORE reports	angiolab random aetrtmt	

to document during program development and to verify and enlarge on the documentation as each program is finalized.

A header similar to that depicted in Figure 15.8 should be placed at the start of each program. If the program is modified, the date and name of the person making the modification should be noted.

```
***Begin check for patients with missing gender
  data ***;
***This section makes a comparison of the two
  adjunct treatment groups ***;
```

Comments similar to these should precede each step in the program, and the time required for documentation (about 10% of the time required for the program itself) should be incorporated in time lines and work assignments. A summary table listing all programs should be maintained as in Table 15.8.

TO LEARN MORE

Abramson NS; Kelsey SF; Dafra P; Sutton-Tyrell KS. (1992). Simpson's paradox and clinical trials: What you find is not necessarily what you prove. *Ann. Emerg. Med.* 21:1480–1482.

Altman DG; Dore CJ. (1990). Randomisation and baseline comparisons in clinical trials. *Lancet.* 335:149–153.

Bailar JC; Moseteller F. (1992). *Medical Uses of Statistics*, 2nd ed. Boston: NEJM Books.

Begg CB. (1990). Suspended judgment: Significance tests of covariance imbalance in clinical trials. *Control. Clin. Trials* 11:223–225.

Berger V. (2000). Pros and cons of permutation tests. *Statist. Med*. 19:1319–1328.

Berger V; Permutt T; and Ivanova A. (1998).The convex hull test for ordered categorical data. *Biometrics* 54:1541–1550.

Cleveland WS. (1984). Graphs in scientific publications. *Am. Statist.* 38:261–269.

Conover W; Salsburg D. (1988). *Biometrics* 44:189–196.

Dar R; Serlin A; Omer H. (1994). Misuse of statistical tests in three decades of psychotherapy research. *J. Consult. Clin. Psychol.* 62:75–82.

Donegani M. (1991). An adaptive and powerful test. *Biometrika* 78:930–933.

Entsuah AR. (1990). Randomization procedures for analyzing clinical trend data with treatment related withdrawals. *Comm. Statist*. A. 19:3859–3880.

Feinstein AR. (1976). Clinical biostatistics XXXVII: Demeaned errors, confidence games, nonplussed minuses, inefficient coefficients, and other statistical disruptions of scientific communication. *Clin. Pharmacol. Ther.* 20:617–631.

Fienberg SE. (1990). Damned lies and statistics: Misrepresentations of honest data. In Editorial Policy Committee, *Ethics and Policy in Scientific Publications.* Council of Biology Editors, pp. 202–206.

Freedman L; Sylvester R; Byar DP. (1989). Using permutation tests and bootstrap confidence limits to analyze repeated events data from clinical trials. *Control. Clin. Trials* 10:129–141.

Gail MH; Tan WY; Piantadosi S. (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika* 75:57–64.

Hilton J. (1996). The appropriateness of the Wilcoxon test in ordinal data. *Statist. Med*. 15:631–645.

Howard M. (pseud for Good P.). (1981). Randomization in the analysis of experiments and clinical trials. *Am. Lab*. 13:98–102.

International Study of Infarct Survival Collaborative Group. (1988). Randomized trial of intravenous streptokinase, oral aspirin, both or neither, among 17,187 cases of suspected acute myocardial infarction. ISIS-2. *Lancet.* 2:349–362.

Lachin JM. (1988). Properties of sample randomization in clinical trials. *Contr. Clin. Trials* 9:312–326.

Lachin JN. (1988). Statistical properties of randomization in clinical trials. *Contr. Clin. Trials* 9:289–311.

Manly B; Francis C. (1999). Analysis of variance by randomization when variances are unequal. *Aust. New Zeal. J. Statist.* 41:411–430.

MacKay RJ; Oldford RW. (2001). Scientific method, statistical method and the speed of light. *Statist. Sci.* 15:254–278.

Mehta CR; Patel NR; Gray R. (1985). On computing an exact confidence interval for the common odds ratio in several 2 × 2 contingency tables. *JASA* 80:969–973.

O'Brien P. (1988). Comparing two samples: extension of the *t*, rank-sum, and log-rank tests. *JASA* 83:52–61.

Oosterhoff. J. (1969). *Combination of One-sided Statistical Tests*. Amsterdam: Mathematisch Centrum.

Peddiwell JA; Benjamin HH (1959). *Saber-Tooth Curriculum*. New York: McGraw-Hill.

Pesarin F. (2001). *Multivariate Permutation Tests*. NewYork: Wiley.

Pledger GW. (1992). Basic statistics: importance of adherence. *J. Clin. Reas. Pharmacoepiemiol* 6: 77–81.

Pothoff RF; Peterson BL; George SL. (2001). Detecting treatment-by-center interaction in multi-center clinical trials. *Statist. Med*. 20:193–213.

Salsburg DS. (1992). *The Use of Restricted Significance Tests in Clinical Trials*. New York: Springer.

Shih JH; Fay MP. (1999). A class of permutation tests for stratified survival distributions. *Biometrics* 55:1156–1161.

Troendle JF; Blair RC; Rumsey D; Moke P. (1997). Parametric and nonparametric tests for the overall comparison of several treatments to a control when treatment is expected to increase variability. *Statist. Med.* 16:2729–2739.

Wears RI. (1994). What is necessary for proof? Is 95% sure unrealistic? (Letter) *JAMA* 271:272.

Weerahandi S. (1995). *Exact Statistical Methods for Data Analysis*. Berlin: Springer.

Wei LJ; Smythe RT; Smith RL. (1986). K-treatment comparisons in clinical trials. *Ann. Math. Statist*. 14:265–274.

Westfall DH; Young SS. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: John Wiley.

Zelen M. (1971). The analysis of several 2 × 2 contingency tables. *Biometrika* 58:129–137.

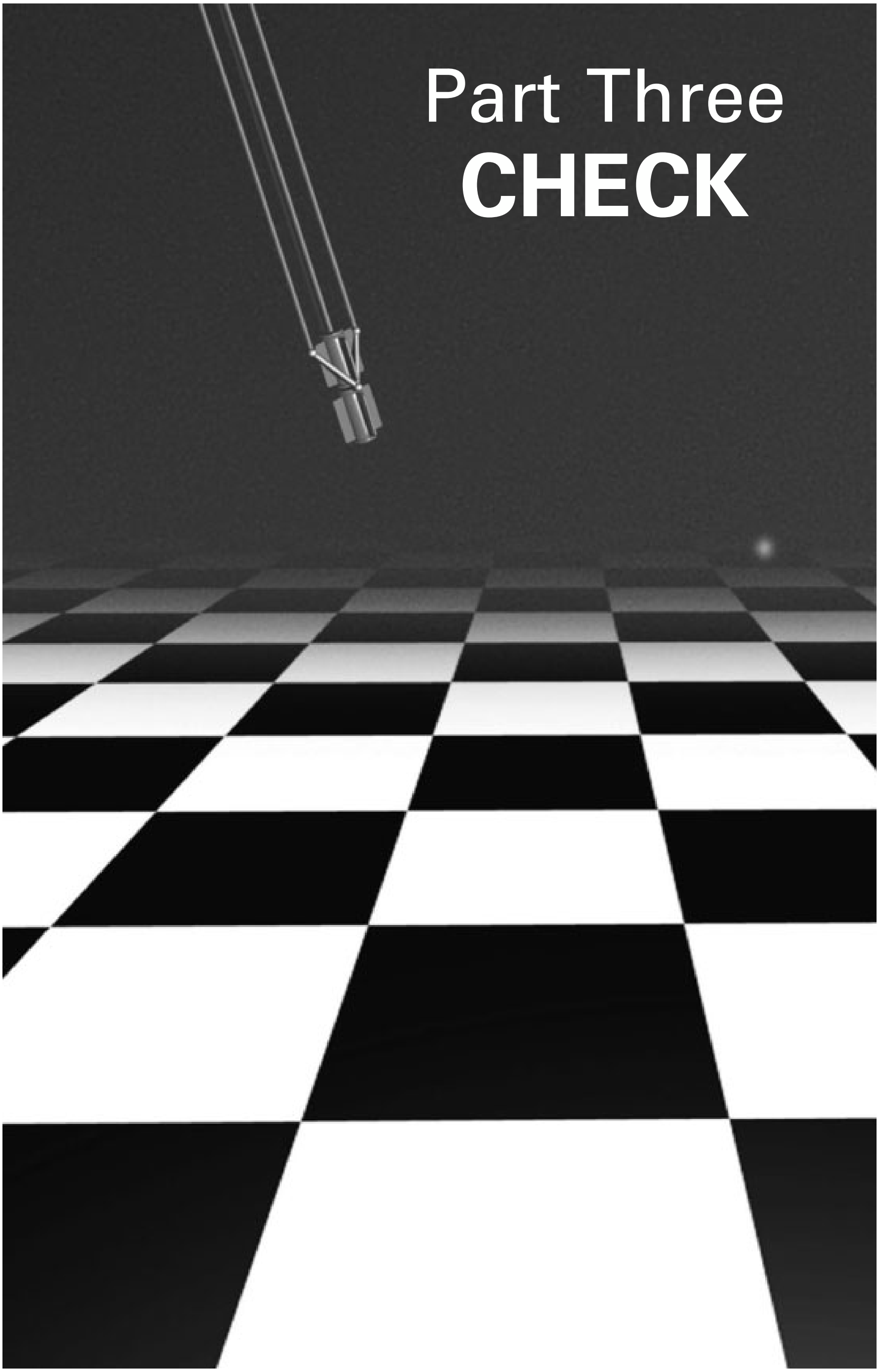
A PRACTICAL GUIDE TO STATISTICAL TERMINOLOGY

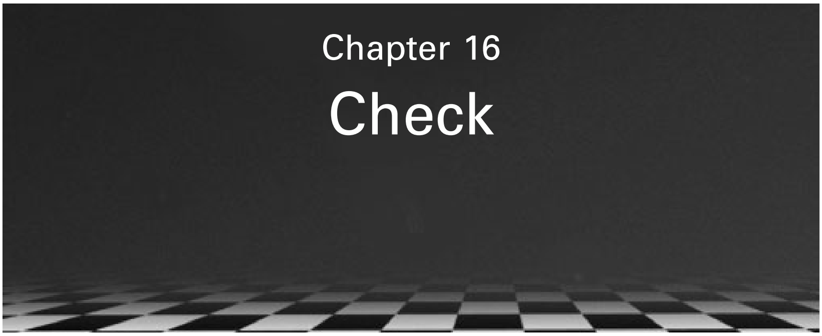
Statisticians tend to utilize their own strange and often incomprehensible language. Here is a practical guide to the more commonly used terms. Italicized terms are included in this glossary.

Analysis of variance	**A technique for analyzing data in which the effects of several factors are taken into account simultaneously such as treatment, sex, and the use of an adjunct treatment. *p-Values* obtained from this technique are often suspect as they rely on seldom realized assumptions.**
Arithmetic mean	**Also known as the arithmetic average or simply as the mean or average. The sum of the observations divided by the number of observations. Can be deceptively large when the distribution is skewed by a few very large observations as in a distribution of incomes or body weights. The *median* should be reported in such cases.**
Chi-square distribution	**A distribution based on theoretical considerations for the square of a normally distributed random variable.**
Chi-square statistic	**A test statistic based on both the observed values in a contingency table and the values one would expect if the *null hypothesis* were true. With tens of thousands of observations, the chi-square statistic will have the *chi-square distribution*. With small samples, it may have a quite different distribution.**
Confidence limits	**The boundary values of a confidence interval.**
Critical value	**That value of a test statistic that separates the values for which we would reject the hypothesis from the values for which we would accept it.**
Exact test	**The calculated *p-value* of the test is exactly the probability of a *Type I error*; it is not an approximation.**
Logistic regression	**A statistical method applied to time-to-event data. Applicable even when observations are censored. Used both to extrapolate into the future and to make treatment comparisons.**
Median	**The 50th percentile. Half the observations are larger than the median and half are smaller. The *arithmetic mean* and the median of a *normal distribution* are the same.**
Minimum relevant difference	**The smallest difference that is of clinical significance.**
Normal distribution	**A symmetric distribution of values that takes a bell-shaped form. Most errors in measurement follow a normal distribution.**

Null hypothesis	**The hypothesis that there are no (or null) differences in the populations being compared.**
Permutation tests	**A family of statistical techniques using a variety of test statistics. The *p-values* obtained from these tests are exact, not approximations.**
***p*-Value**	**The probability of observing by chance alone a value of the statistic more extreme than the observed value.**
Rank tests	**The ranks of the observations are used in place of their original values to diminish the effects of extreme observations.**
Significance level	**Probability of making a *Type I error*.**
Student's *t*	**See *t-test*.**
***t*-Test**	**A technique utilizing the *Student's t* statistic for comparing the means of two samples.**
Type I error	**Attributing a purely chance effect to some other cause.**
Type II error	**Attributing a real effect to chance.**
Wilcoxon test	**Like the *t-test*, compares the means of two samples, but uses ranks rather than the original observations.**

Part Three
CHECK





Chapter 16

Check

YOU MAY HAVE TURNED in your report to the regulatory agency. They may even have granted the approval you desired. But unless your company plans on going out of business tomorrow, you still have six important issues to resolve.

1. **How will you bring closure to the trials—parting with patients, archiving the data, and publishing the results?**
2. **What did the trials really cost?**
3. **What have you learned during the investigation that would guide you in expanding or narrowing your original claim?**
4. **Are there potential adverse effects that warrant further investigation?**
5. **What have you learned about other diseases and devices or medications that might be of interest to your company?**
6. **What have you learned that would help you to conduct more effective studies in the future?**

CLOSURE

Trial closure has four important aspects: providing for follow-up patient care, making arrangements for storing the data, arranging for publication of the results, and, if the product is brought to market, providing for post-market surveillance.

Patient Care

A patient cannot be discharged from the study until arrangements have been made for continued medical care either from the patient's regular physician (at the patient's expense) or from the appropriate public agency.

If the new treatment represents a demonstrated improvement over existing methodologies, continued supplies must be made available to the patient at no cost until marketing approval is obtained from the regulatory agency.

If the treatment requires a tapering-off phase (e.g., as do beta-blockers), then supplies must be made available to the each patient until a transition to an alternate treatment is complete.[51]

Data

The original case report forms should be stored in a readily retrievable form. (An e-submission will automatically fill this need.) Copies of the master database should be kept both on and off site, at least initially. With the examples of diethylstilbesterol and silicon implants before us, and an increasingly litigious society, it is best to plan on an indefinite storage period for at least one of the copies.

Maintaining archives for samples, X rays, angiograms, and analog EKG and EEG traces can be somewhat more challenging but is also essential. See Bell, Curb, and Friedman (1985).

Spreading the News

Klimt and Canner (1979) recommend that in disclosing the results of the trials you follow this sequence: (1) investigators, (2) participants and their physicians, (3) the scientific press, and (4) marketing materials. See also the *AMA Manual of Style* (1994), Bailar and Mosteller (1988), and Long and Secic (1997).

POSTMARKET SURVEILLANCE

Ours is a litigious society. You want to remain aware of any adverse events that could be attributed—rightly or wrongly—to your product or process. Designate an individual (or department) to handle post-market review; provide them with an 800 number and email address. Encourage physicians to report all unanticipated responses to your product, favorable or unfavorable, to this individual. Pay particular attention to adverse events that come to light during your post-trial review.

BUDGET

Considering that pharmaceutical and device firms, large or small, are by definition profit-making concerns, it is amazing how few investigator (none in my experience) ever bother to complete a post-trial review of the trial budget. Alas, those who do not learn from the lessons of history will be forced to repeat them. You cannot control costs nor spend your money efficiently until you know what your expenses are.

---

[51] See, for example, Bell, Curb, and Friedman (1985).

Your primary objective is to determine the cost of the trials on a per patient basis. Your secondary objectives are to determine the impact of any actual and potential cost cutting.

Variable and Fixed Expenditures

The cost per patient can be divided into variable and fixed costs. Variable costs include costs of hospitalization (if any), physician visits, drugs and devices, special procedures (angiograms, EKGs) and any other miscellaneous costs *that can be attributed to* a specific patient.

Fixed costs include work-hours invested by you and your staff on all phases of the trials, computer hardware and software, travel, and all other costs that cannot be attributed to patients *whose results were used to determine the effectiveness of treatment*. In other words, costs associated with patients who were interviewed but declared ineligible, who dropped out along the way, and whose records are incomplete should also be attributed to fixed costs.

CONTROLLING EXPENDITURES

You knew at the start that the most effective way to control costs (apart from the switch from printed forms to electronic data capture) was to hire the right investigators and closely monitor their efforts, and to recruit only those patients who would make a positive contribution to the trials. Of course, this goal is seldom achieved. Now is the time to document anything you have learned during the trials that will help you come closer to this goal on the next go around.

Inevitably, at one or more points during a set of lengthy trials, a decision is made to trim costs. Not infrequently the decision is external to the trials themselves, a corporate-level decision, but you as a middle manager had no choice but to go along. Did you make cuts in the appropriate places? What costs ought you to have trimmed instead?

If you'd had more money to spend, how would you have spent it?

TRIAL REVIEW COMMITTEE

The majority of the remaining issues are best resolved with the aid of a post-trial committee or committees. Membership should include all

SEND THE MANAGERS OUT OF THE ROOM

Recently I was called into help with a health survey that had gone awry. A 70% response rate with the pilot had fallen to 35%. The discussion went round and round with all present testifying that the principle survey had been virtually identical with the pilot. What to do? The manager was called away by a family emergency at that point, and I was asked to close out the meeting. Within minutes of her departure, horror story after horror story was elicited as the changes and compromises that had gone into the principal survey were revealed. We soon documented a list of positive suggestions and the survey was back on track. (Not incidentally, some of the compromises had been the result of miscommunication and a failure to develop a shared understanding of objectives.)

The bottom line. Team members are often reluctant to speak up for fear they may be blamed or viewed as criticizing the boss. A third-party facilitator should be called in to lead a group discussion of the design and implementation teams *without the managers present*.

the original members of the design committee if available, representatives from the implementation team (which representatives will depend on the issues that have arisen during the trials), all the investigators, one or more members of the safety committee, and representatives of *all other* project teams in your company. The concepts you are about to review will be of value throughout the organization.

I'd recommend, though, that the members of the design team, the CRMs, and medical monitor meet separately with the investigators.

AFTER ACTION REVIEW

The questions to ask will depend on whether the new treatment proved to be a success or a failure and whether the trials themselves were conclusive.

When the treatment is a success, you need to ask:

- **Was the treatment more effective with some groups of patients than with others?**
- **Were interactions with other treatments uncovered?**
- **Are there other groups of patients and other indications to which the trials could be extended?**
- **Are there indications of possible long-term adverse effects? You will want to pay particular attention to these during postmarketing surveillance.**

When the treatment is a success, it is not difficult to obtain suggestions for a second set of trials involving a new category of patients or a new indication. But when the treatment is a failure, it is sometimes difficult to find anyone after the trials are over who is knowledgeable enough to suggest a cause. As a new treatment's inadequacies becomes evident, people leave a project, are reassigned, or find reasons for leaving. An immediate postmortem of the trial findings is even more important in such circumstances.

Sometimes the reason for failure is evident. A rare but serious side effect was brought to light. Sometimes it can be subtler: Too small a starting sample accompanied by a larger than anticipated number of withdrawals led to an inconclusive finish.

Often the principal reasons for failure are those that should have discouraged the trials launch in the first place: too little was known about the mode of action, about possible interactions with other commonly prescribed treatments. A surgical procedure still was evolving, and consequently took on a dozen different forms in the hands of the various investigators.

A treatment can sometimes appear to be a failure solely because the results were inconclusive. Sufficient patients were not enrolled in the trials, and to the point, too few patients remained after correcting for dropouts, treatment withdrawals, and missing observations.

It is important under these circumstances to recompute the power of the test using one of the same computer programs that was used to determine sample size.[52] Excessively low power suggests that additional trials might be called for.[53]

The reasons underlying excessive numbers of withdrawals and dropouts should be investigated. The treatment regimen should undergo a detailed critical examination. (Was the dose too high or too low? Was the treatment regimen so complex as to preclude adherence?)

Whatever the outcome of the trials, additional consideration may need to be given to secondary responses, interactions, and adverse events. Detection of secondary beneficial responses that were unanticipated during the design stage (e.g., a sedative or an antifebrile effect in addition to an anti-inflammatory) can often be gleaned from the anecdotal accounts of investigators. This is a major reason for including all the investigators in the trial postmortem.

Such accounts may suggest a basis for future trials or for future preclinical investigation.

---

[52] See Chapter 6 and the Appendix.

[53] See Barnard (1990) and Young, Bresnitz, and Strom (1983).

Interactions

Breaking down the data into subsets (female patients, patients with ostial lesions) and analyzing each subset separately often reveals both those patients who are most likely to benefit from the treatment as well as those who are least likely. A new treatment that is at first judged to be unsuccessful may actually prove to have demonstrated potential in a particular category of individuals.

Adverse Events

Adverse effects of treatment may remain undetected during clinical trials for at least three reasons:

1. **Restricted set of participants**
2. **Limited follow-up**
3. **Inadequate sample size (particularly for relatively rare side effects)**

Eligibility criteria may have limited the original trials to males under 60 not suffering from collateral conditions. But physicians are always tempted to extend the range of application of a successful treatment regardless of what it might say on the package insert. To avoid potential litigation, you should consider engaging in follow-up clinical trials that would focus on individuals who were excluded from the original study.

Additional, long-term monitoring of at least a subset of those patients for whom you already have such a great store of data is recommended. It offers a way of discovering both (1) additional beneficial effects and (2) delayed adverse effects (thus avoiding or minimizing the impact of subsequent litigation).

Postmarketing analysis typically is based solely on anecdotal accounts. Systematic follow-up has a far greater probability of early detection and countering false claims.

For example, while many cholesterol-lowering drugs are now on the market, as of this writing only one had sufficient carefully monitored post-trial experience that its manufacturers were permitted to make the claim that it reduces coronary mortality. Naturally this drug now has the largest market share. And, since I have an adverse reaction to the drug in question—my physician's and my first choice—I can only hope my present lipid-lowering drug will ultimately prove equally effective.

Note. In most countries, you will need to obtain permission from patients to continue surveillance after the scheduled end of the trials.

PROJECT WISE, CORPORATION STUPID

Bumbling Pharmaceuticals had a well-established market share for its *A* device, but was hoping that *E*, its new experimental aid, might give it total market dominance. Alas, an initial analysis of the data showed that *E* offered no particular advantage over *A*. Probing deeper, it was found that when an adjunct given some of the patients was taken into account, patients without the adjunct did do better with *E* in some cases. The adjunct helped patients implanted with the traditional device *A* but only worsened their condition when used in conjunction with *E*.

A further analysis on the basis of sex revealed that while almost 50% of the women given the adjunct along with the experimental procedure suffered a relapse, the standard procedure given in conjunction with the adjunct was 100% effective.

One could visualize the headline, "Bumbling Device 100% effective in women when used in accordance with doctor's instructions," and the resultant improvement in Bumbling's bottom line. But then, I was thinking from the corporation's point of view. In the end, the marketing representative decreed we would just report the combined results. "Our job," she said, "is only to report on the new device."

COLLATERAL STUDIES

I cannot stress sufficiently the importance of including representatives of all study teams, past and present, on post-trial committees. Every trial results in the uncovering of information that may prove of value in collateral studies or suggests additional applications for existing products. Vice versa, a phenomenon that was not well understood in your own trials, and may have proved a barrier to their effective completion, may already have been encountered and overcome by another study team.

Many companies today allow employees to pursue MBAs or doctorates on company time. Yet nothing could be more valuable to an employee and less costly to the company than the shared experiences of other employees.

FUTURE STUDIES

As an aid to future investigations you need to answer and document your answers to all of the following questions:

Data

- **Were essential baseline variables and risk factors neglected?**
- **Were there baseline imbalances? How could they have been prevented?**

- **Was blinding maintained? If not, why not?**
- **Did you gather all the information you needed? What other observations should you have recorded? Should alternative measuring techniques have been used?**
- **Did you gather redundant information? What information should have been omitted? Would there be associated cost savings?**

Patients

- **Which recruiting strategies were the most effective?**
- **What was the time course of recruitment?**
- **How could the ratio of eligible to ineligible patients be increased?**
- **Were some sites more effective at retaining patients (minimizing withdrawals) than others? Why were they successful?**
- **Which investigators should be asked to participate in future trials?**

In short, the function of post-trial review is to elicit and document anything and everything that might be of assistance to you and your coworkers in future efforts. See also Beck (1996).

TO LEARN MORE

American Medical Association. (1994). *AMA Manual of Style*.

Bailar JC III; Mosteller F. (1988). Guidelines for statistical reporting for articles in medical journals: Amplifications and explanations. *Ann. Intern. Med.* 108:266–273.

Barnard GA. (1990). Must clinical trials be large? The interpretation of *p*-values and the combination of test results. *Statist. Med.* 9:601–614.

Beck B. (1996). *Clinical Trials: Decision Tools for Measuring and Improving Performance*. Buffalo Grove, IL: Interpharm Press.

Bell RL; Curb JD; Friedman LM, et al. (1985). Termination of clinical trials: Beta-blocker heart attack trial and the hypertension detection and follow-up program experiences. *Control. Clin. Trials* 6:102–111.

Klimt CR; Canner PL. (1979). Terminating a long-term clinical trial. *Clin. Pharmacol. Ther*. 25:641–646.

Long TA; Secic M. (1997). *How to Report Statistics in Medicine.* Philadelphia: American College of Physicians.

Schulz KF; Chalmers I; Grimes DA; Altman DG. (1994). Assessing the quality of randomization from reports of controlled trials published in journals of obstetrics and gynecology. *JAMA* 272:125–128.

Schulz KF; Chalmers I; Hayes RJ; Altman DG. (1995). Empirical evidence of bias: Dimensions of methodological quality associates with estimates of treatment effects in controlled trials. *JAMA* 273:408–142.

Young MJ; Bresnitz EA; Strom BI. (1983). Sample size nomograms for interpreting negative clinical studies. *Ann. Intern. Med*. 99:248–251.

Appendix

Software

Let the software determine your hardware.

Good (1984)

An extremely wide choice of software is available to ease your task of designing, managing, and analyzing clinical trials. Some programs offer to do it all, while others, more specialized, provide for speech recognition and collecting information from handheld devices.

CHOICES

You will need at least five types of programs; whether you buy them separately or in a comprehensive integrated package is up to you.

1. **Project management**
2. **Data entry**
3. **Data management**
4. **Data analysis**
5. **Utilities**

All in One

TrialXS. Fully integrates trial management, electronic data capture and data management into a single environment. Trial XS/TMS. ClinSource USA, 2678 Bishop Drive, San Ramon, CA 94583; ClinSource NV, Mechelsesteenweg 455 Bus 2, B-1950 Kraainem, Belgium, + 32 (0)2-766-00-80, *info@clinsource.com*.

Oracle Clinical. Has the standard features of Oracle. Automatically create views corresponding to each case report form (CRF) and automatically extracts data into SAS for analysis. It's easy to create custom views combining data from multiple CRFs, to query the data on line and to create any number of data snapshots for interim analysis during normal data processing.

Provides your staff with the ability to visualize the planned, projected, and actual patient enrollment and study time lines, develop detailed visit schedule specification and tracking, including the identification of missing and late CRFs, manage and track treatment blind breaks, and track patient availability and withdrawal information.

On the down side, Oracle Clinical is less flexible than Oracle and we've found it easier to create reports and CRFs in the original, less expensive product. Oracle Corporation, 1-888-672-2534.

Project Management

TrialWorks™. Tracks your project management data by study and by site. Multiple users can simultaneously access all your tracking data. Tracking is comprehensive and includes IRB approvals, 1572s, patient enrollment and withdrawals, regulatory submissions, and investigator, vendor, and CRA payments. TrialWorks then uses those data to produce over 100 reports. TrialTrac, Inc. 7336 Santa Monica Boulevard, W. Hollywood, CA 90046, (323) 930-9842. *www.trialtrac.com*.

Almost All in One

SAS. SAS afficinados swear that with all modules in place, SAS can be used to build e-CRFs, manage the data, and perform the analyses. We're not one, so we've listed SAS below under "Data Analysis."

Clindex. Data entry, data management, and payment management. Easy-to-build screens can be generated directly from any word-processor document. Then add edit checks and data validation, radio buttons, and drop-down menus. Built around the Sybase InfoMaker® screen and report painter and Sybase SQL Anywhere® database, Clindex provides an easy to use SQL environment. Double clicking on a patient or CRF in any report can be used to display all that patient's CRFs in read-only or update mode. Fortress Medical Systems, Inc., 310 7th Avenue North Hopkins, MN 55343, (952) 988-9123, *mjones@fortressmedical.com*.

DATA ENTRY

Handheld Devices

Gather information from the patient's beside or have the patient himself record the exact time and date medication was taken or symptoms observed.

Touch Screen

HCOL Clinical Study. Includes electronic touch screen and web interfaces for many commonly used patient questionnaires. Spinal Outcomes Lumbar (SOL), Spinal Outcomes Cervical (SOC), Electronic Pain Diagram, Visual Analog Scale (VAS), Medical History Questionnaire, SF-36, the Oswestry Disability Questionnaire, Neck Disability Index, Johns Hopkins Cervical and Lumbar Outcomes Questionnaires, Others . . . , *http://www.hcol.com/*.

Speech Recognition

PocketTrials. Uses structured speech recognition to make it easy to enter data in hands-busy environments. Support is catch as catch can. (301) 776-1196, *support@PocketTrials.com*.

e-CRFs

When evaluating products in this category, look for the following essentials:

- **Range and logic validation checks**
- **Pull-down selection menus**
- **Radio buttons and toggle buttons**

Do It Yourself

StudyBuilder. Even if you decide to buy another, more expensive product, Study-Builder is the ideal way to introduce your CRMs and medical staff to computerized case report forms. The system utilizes a point and click method that allows you to build a form by dragging questions out of their extensive built-in library of validated study questions, and dropping them into place. Fields are validated and pop-up warning messages appear on the screen if bounds are exceeded.

Build the form in a language with which your staff is familiar, then have it immediately translated into Dutch, English, French, German, Italian, Spanish, Modern Standard Arabic, or Japanese for use in other countries. Contains built-in support for downloading data onto PCs via serial, infrared and USB ports on many platforms including the Web! 268 Bush Street, Suite 1123, San Francisco, CA 94104, *us@studybuilder.com*; Postbus 177, 1000 Ad Amsterdam, The Netherlands, +31 (71) 514-2988, nl@studybuilder.com; Level 11 Park West Building, 6-12-1 Nishi-Shinjuku, Shinjuku-Ku Tokyo, 160-0023 Japan, *jp@studybuilder.com*.

Sybase InfoMaker®. Screen and report painter lets you build the most user-friendly forms. Requires sophisticated programmers. Accepts PowerBuilder input. For Windows, UNIX, or LINUX. Sybase, 6475 Christie Avenue Emeryville, CA 94608, (510)922-3500.

Via Web Access

Clintrinet. Each trial is assigned a Web site that becomes the central workplace for all trial personnel, and warehouse for all trial data and records. Data entry includes range and logic checks. ClinicalTrialsNet Inc., 12 John Street, Charleston, SC 29403, (877) 552-8638, *info@clinicaltrialsnet.com*.

Datatrak EDC™. Am Propsthof 80, 53121 Bonn, Germany, +49-228-979-8330, *products@datatraknet.com*.

MetaTrial System. Clinical data are entered into the RDE module installed at remote study sites where data are validated and transmitted to the central site via a secure e-mail system. The central site module collects the incoming data and translates the data into a SAS compatible file. 2001 CB Technologies, Inc., 350 Eagleview Boulevard, Exton, PA 19341 (610) 280-7400.

DATA MANAGEMENT

When selecting a database management system, there are five key areas on which to focus: (1) ease of formulating queries, (2) speed of retrieval of data, (3) ease of updating data, (4) ease of restructuring the database, (5) ease of integrating other applications including data entry and statistical software.

Oracle. Comprehensive and reliable. If you think like a programmer, then Oracle is remarkably easy to use to create or interrogate a clinical database. For UNIX Oracle Corporation, 1-888-672-2534, *http://www.oracle.com/global/index.html?content.html*.

Sybase SQL Anywhere. For Windows, UNIX, or LINUX. Sybase, 6475 Christie Avenue, Emeryville, CA 94608, (510)922-3500.

C-ISAM. Although not a relational database, you don't have to sift through records to get to the data you want. B+ tree index architecture makes data retrieval fast and easy. C-ISAM uses index entries as keys that point to records. These keys allow you to find the specific pieces of data you want, without having to look at extra records. On top of that, C-ISAM uses techniques to compress the keys for efficient index storage and processing. The reduced key size means faster response and better performance for the end user. For UNIX: IBM Data Management Sales, 16011 College Boulevard, Lneexa, KS 66219, *moreinfo@informix.com*.

DATA ENTRY AND DATA MANAGEMENT

Small-Scale Clinical Studies

Microsoft Access. Choice of spreadsheet entry or more sophisticated forms. Provides toggle buttons but no pull-down menus. Includes range and logic checks. PC-based and can be purchased from virtually any computer or office supply outlet.

Advanced Revelation. Excellent data manager with minimal memory requirements. Flexible data entry, including pull-down menus, range and logic checks. Revelation Software 800/262-4747, + 44 (0) 1908-233255, +61 -2-9939-6399, *info@revelation.com*.

Paradox. SQL-like interface with spreadsheet views. Range and logic checks, but lacks ability to create other user-friendly features. Corel Corporation, 1600 Carling Avenue, Ottawa, Ontario, Canada K1Z 8R7.

Clinical Database Managers

The members of this extensive class are both expensive, and generally unsatisfactory because they force you to adapt your study to their software, thus violating the first rule of trial design to let your reports determine the data to be collected.

Acceliant™. Enmed, 35 Crosby Drive, Suite 201, Bedford, MA 01730, (781) 999-1000, *enmedinfo@enmed.com*.

ACS Study Manager. Clinsoft Corporation, 10 Maguire Road, Lexington, MA 02421, 1.800.551.4440, *www.clinsoft.net*.

Clintrial. This product has seen unsettling times during a period of mergers and acquisitions, but emerged looking better than ever. Worth a second look. Phase Forward Incorporated, 1440 Main Street, Waltham, MA 02451-1623 +1-888-703-1122.

SyMetric. SyMetric Sciences, Inc. 1-2082 Sherbrooke West, Montreal, Quebec. Canada H3H 1G5, Major@SyMetric.ca or *http://www.symetric.ca/*.

DATA ANALYSIS

SAS. Over-priced, cumbersome, and unevenly documented. Large number of statistical routines with many options for table creation and graphs. Too few built-in nonparametric routines, but the statistical literature is filled with SAS macros for a wide variety of supplemental procedures including bootstrap and density estimates. Knowledgeable programmers are essential but widely available.

Thoroughly validated and has been used in hundreds of submissions. SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513 (Windows, MVS, CMS, VMS, Unix), *www.sas.com*.

SPSS. The poor person's SAS. Offers ease of use along with a large number of statistics. A bootstrap subcommand provides bootstrap estimates of standard errors and confidence limits. Thoroughly validated and has been used in dozens of submissions (Windows). SPSS, Inc., 444 North Michigan Avenue, Chicago, IL 60611, 312-329-2400, *www.spss.com*.

Stata. Provides a comprehensive set of statistics routines plus subroutines and pre-programmed macros for bootstrap, density estimation, and permutation tests. Programmable with many flexible graphics routines (Windows, Unix). Stata Corp, 702 University Drive East, College Station, TX 77840, 1-800-782-8272, *www.stata.com*.

Data Desk/Activ Stats/DataDesk XL. The best program I know for exploratory data analysis. Windows and Macintosh versions, *info@datadesk.com* or *http://www.datadesk.com/DataDesk/system.shtml*.

StatXact. While not a comprehensive statistics package, it is a must for the exact analysis of contingency tables (categorical or ordered data) and should be purchased along with one of the four statistics programs listed above. The StatXact manual is a textbook in its own right. The program is thoroughly validated and has been used in dozens of submissions. Versions for Windows or Unix. Also available as an add-on module for both SAS and SPSS. Cytel Software Corporation, 675 Massachusetts Avenue, Cambridge, MA 02139, (617)661-2011, *www.cytel.com*.

LogXact. Supports inference for logistic regression and provides conditional logistic regression for matched case-control studies or clustered binomial data. Cytel Software Corporation, 675 Massachusetts Avenue, Cambridge, MA 02139, (617)661-2011, *www.cytel.com*.

NPC TEST. The only statistics program on the market today that provides for multifactor analysis by permutation means. Cutting edge, but has yet to be validated. A demonstration version, SAS macro, and S-Plus code may be downloaded from *http://www.stat.unipd.it/~pesarin/software.html*.

UTILITIES

For Sample Size Determination

PASS 2000. Lets you solve for power, sample size, effect size, and alpha level and automatically creates appropriate tables and charts of the results. Covers an extremely wide range of statistical procedures including Fisher's Exact Test, the Wilcoxon test, factorials, and repeated measures. The 490-page manual contains tutorials, examples, annotated output, references, formulas, verification, and complete instructions on each procedure. NCSS Statistical Software, 329 North 1000 East, Kaysville, UT 84037, (800) 898-6109 Download free demo version from *http://www.ncss.com*.

nQuery Advisor. Helps you determine sample size for 50+ design and analysis combinations (Windows). Statistical Solutions, 8 South Bank, Crosse's Green, Cork, Ireland, 800/262-1171, +353 21 4319629, *www.statsol.ie*.

Screen-Capture

Winrunner. Mercury Interactive, 1325 Borregas Avenue, Sunnyvale, CA 94089. 1-800-TEST-911, +33 1 40 83 68 68, *Info-France@mercury-eur.com*.

Silktest. Segue Software, 201 Spring Street, Lexington, MA 02421, 1-800-287-1329, +49-211-9943 605, +81(0)3-3279-0771, *info@segue.com*.

Data Conversion

DBMS Copy. Lets you exchange files among two dozen statistics packages, a dozen plus data base managers, and multiple versions of a half-dozen spreadsheets. UNIX and Windows versions. Conceptual Software, 9660 Hillcroft, suite 510, Houston, TX 77096, 1-800-328-2686, *www.conceptual.com*.

Author Index

Abramson, N. S., 195
Adams, J. R., 63
Ader, H. J., 64
Agras, W. S., 105
Altman, D. G., 180, 195, 208
American Medical Association, 202, 208
Angell, M., 6, 50, 54
Applegate, W. B., 105
Armitage, P., 161, 164
Asilomar Working Group on Recommendations for Reporting Clinical Trials in the Biomedical Literature, 94

Backhouse, M. E., 63
Bailar, J. C., 195, 202, 208
Baker, A., 106
Barbui, C., 49, 63
Barkhof, F., 64
Barnard, G. A., 205, 208
Baron, J., 106
Barr, R. G., 106
Barsky, A. J., 106
Beck, B., 208
Begg, C. B., 94, 195
Bell, R. L., 202, 208
Benjamin, H. H., 189, 196
Bennett, C. L., 63
Berger, V., 174, 195
Berlin, J. A., 63
Beveridge, R. A., 46
Birnbaum, Y., 45
Blair, R. C., 196
Blaskowski, T. P., 106
Bluman, L. G., 107
Bradiford, R. H., 105
Bresnitz, E. A., 205, 208
Butow, P. N., 105
Byar, D. P., 195

Campbell, M., 63, 64
Canner, P. L., 202, 208
Cantor, A., 63
Carey, T. S., 96, 105
Cassileth, B., 64
CAST (Cardiac Arrhythmia Suppression Trial), 33, 45
Cavan, B. N., 63
Celano, P., 63
Chalmers, I., 208
Chalmers, T. C., 52, 63
Cheng, S. C., 64
Cho, M., 94
Chow S.-C., 44
Claxton, K., 63
Cleveland, W. S., 195
Cocchetto, D. M., 44
Coglianese, M. E., 106
Conover, W., 185, 195
Conway, M. D., 45
Cramer, J. A., 105, 106
Crowley, J., 46
Curb, J. D., 100, 106, 202, 208
Curtis, R. C., 105
Dafra, P., 195
Dar, R., 195
Date, C. J., 140
Davis, C. E., 104, 105
DeMets, D. L., 40, 44, 45, 99, 104, 162, 164
Desmond, J., 45
Djulbegovic, B., 49, 63
Donegani, M., 195
Dore, C. J., 180, 195
Dowsett, S. M., 105

Eastwood, S., 94
Ebi, O., 45
Ederer, F., 52, 63
Ellis, P. M., 100, 105
Elwood, J. M., 63
Entsuah, A. R., 195
Ernst, E., 64
Evans, M., 104, 106
Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 94. *See also* ICH

Favalli, G., 42, 45
Fay, M. P., 196
Fazzari, M., 30, 45

Feinstein, A. R., 43, 104, 106, 195
Fields, K. K., 63
Fienberg, S. E., 195
Fleming, T. R., 33, 40, 45, 164
Francis, C., 184, 196
Freedman, L., 13 , 188, 195
Friedman, L. M., 44, 99, 104, 202, 208
Furberg, C. D., 44, 99, 104

Gail, M. H., 195
Gandy, B. G., 139, 140
Garattini, S., 63
Garcia-Molina, H., 140
George, S. L., 46, 196
Gillum, R. F., 106
Givens, S. V., 94
Goldman, D. P., 45
Good, P. I., 48, 80, 94, 173, 177, 188, 191, 192, 195
Gordon, D. J., 105
Gordon, M. E., 106
Grambsch, P. M., 56, 64
Gray, R., 196
Greene, H. L., 42, 46
Grimes, D. A., 208

Haidich, A. B., 106
Hamrell, M. R., 164
Handberg-Thurmond, E., 101, 106
Harrington, D., 40, 46
Harris, R., 105
Hayes, R. J., 208
Haynes, R. B., 102, 106
Heller, G., 30, 45
Herman, A. A., 46
Hilton, J., 184, 195
Holborow, D. W., 106
Horton, R., 94
Howard, M., 195
Hujoel, P. P., 106
Hunninghake, D. B., 106

Iber, F. L., 45
ICH, 63, 91, 94
International Committee of Medical Journal Editors, 94
International Study of Infarct Survival Collaborative Group, 190, 195
Ioannidis, J. P., 106
Ivanova, A., 161, 164, 174, 195

Jones, B., 63

Katz, R. J., 46
Keith, S. J., 38, 46, 106
Kelsey, S. F., 195
Kenward, M. G., 63
Kertes, P. J., 45
Kessler, D. A., 94
Keyserling, T., 105
Kinsinger, L., 105
Klimt, C. R., 202, 208
Kloner, R. A., 45
Knipschild, P., 104, 106
Kuderer, N. M., 63

Lacevic, M., 63
Lachin, J. M., 161, 164, 195
Lan, G., 162, 164
Lang, J. M., 104, 106, 179, 182
Laska, E. M., 45
Leffers, P., 104, 106
Lin, D. Y., 164
Liu, J.-P., 44
Lock, S. P., 98, 106
Long, T. A., 94, 202, T. A., 208
Lung Health Study Research Group, 106
Lyman, G. H., 63

MacKay, R. J., 193, 196
Manly, B. F. J., 56, 64, 184, 196
Margitic, S., 101–102, 106
Marshall, G. D., 105
Maschio, G., 46
Massoth, K. M., 106
Matthews, J. N. S., 64
Matts, J. P., 164
Mattson, M. E., 100, 106
Max, M. B., 45
McArdle R., 100, 106
McBride, P. E., 96, 106
McCormick, M., 105
McInnes, P. M., 45
Mehta, C. R., 162, 164, 173, 196
Mendes, A., 139, 140
Milgrom, P. M., 104, 106
Miller, D. H., 64
Moher, D., 94
Moke, P., 196
Mosteller, F., 195, 202, 208
Moye, L. A., 64
Mulay, M., 45
Murray, P. J., 45

Nardi, R. V., 44
National Cancer Institute, 45
Ness, R. B., 63
Nielsen, O. S., 124

O’Brien, P., 184, 196
Ockene, J. K., 102, 106
O’Connor, M., 94
Oldford, R. W., 193, 196
Oldrizzi, L., 46
Olkin, I., 94
Omer, H., 195
Oosterhoff J., 196

Pablos-Mendez, A., 106
Pajak, T., 46
Paoletti, L. C., 45
Park, T. S., 164
Parker, B., 107
Patel, N. R., 164, 173, 196
Pecorelli, S., 45
Peddiwell, J. A., 189, 196
Permutt, T., 174, 195
Pesarin, F., 163, 164, 184, 196
Peterson, B. L., 196
Piantadosi, S., 64, 195
Pitt, B., 45
Pledger, G. W., 182
Pocock, S., 45
Portenoy, R. K., 45
Posnett, J., 63
Pothoff, R. F., 196
Prien, R. F., 45
Prokscha, S., 140

Redmond, C., 46
Regan, K., 107
Renard, J., 45
Rifkin, R. M., 46
Riley, W. A., 45
Robinson, D. S., 45
Roden, D. M., 46
Rondel, R. K., 140
Rosenberger, W. F., 161, 164
Ruffin, M. T., 106
Rumsey, D., 196
Rush, H., 107
Rutman, O., 94

Sackett, D. L., 102, 106
Sacks, H. S., 63
Salsburg, D. S., 175, 185, 195, 196
Schechtman, K. B., 106
Scher, H. I., 30, 45
Schildkraut, J., 107
Schron, E. B., 102, 106
Schulz, K. F., 208
Schumaker, S. A., 106
Secic, M., 94, 179, 182, 202, 208
Senchaudhuri, P., 164
Serlin, A., 195
Shea, S., 106
Shih, J. H., 196
Shorack, M. A., 100, 107
Schumaker, S. A., 102, 106
Shuster, J. J., 59, 64
Siegmund, H., 162, 164
Simon, R., 64
Slud, E., 162, 164
Smith, H., 63
Smith, R. L., 196
Smythe, R. T., 164, 196
Spilker, B., 45, 97, 106
Stewart, F. M., 46
Strom, B. I., 205, 208
Sturdee, D. W., 102, 107
Sugarman, J., 101, 107
Sutton-Tyrell, K. S., 195
Switula, D., 94
Sylvester, R., 195

Tan, W. Y., 195
Tattersall, M. H., 105
Taylor, D. W., 102, 106
Thall, P. F., 64
Therasse, P., 124
Theriault, R. L., 46
Therneau, T. M., 56, 64
Thompson, A. J., 64
Tilley, B. C., 100, 107
Torgerson , D., 63, 64
Trivedi , M. H., 107
Troendle, J. F., 196
Tsiatis, A. A., 164, 173
Tubridy, N., 56, 64
Ullman, J. D., 140

van Oosterom, A. T., 45, 124
Vantongelen, K., 45
Varley, S. A., 140
Vermorken, J. B., 45
Verweij, J., 124
Vickers, A., 52, 64
Vickers, A., 52, 64
Violante, A., 63
Voodford, F. P., 94

Wears, R. I., 196
Webb, C., 140
Weerahandi, S., 184, 196
Wei, L. J., 42, 62, 64, 161, 162, 164, 196
Weinstein, P., 106
Weiss, R. B., 37, 46
Wells, F., 98, 107
Westfall, D. H., 196
Widom, J., 140
Wieand, H. S., 46
Williams, L. A., 46

Yao, Q., 62, 64
Young, M. J., 205, 208
Young, S. S., 196

Zelen, M., 180, 196

Subject Index

AAR. *See* After-action review (AAR)
Abnormal values, correcting, 12
Academic physicians, 96
Acceliant^{TM} software, 213
Access privileges, 138–139
ACS Study Manager software, 213
Active (positive) controls, 49
Adaptive design, 161
Ad hoc hypotheses, 191–192
Adjuvant treatment, 172
Advanced Revelation software, 213
Adverse effects list, 10
Adverse-event forms, 66–67
Adverse-event reports, 87–88
Adverse events, 24, 34, 59, 103–104, 158–159, 183, 192, 206
  monitoring and reporting on, 66–67, 80
  nature and frequency of, 92
  tracking, 68
Adverse-events policy, 67
Adverse reactions, 66
  methods for handling, 79
Advertising campaigns, 100
After-action review (AAR), 75, 204–207
Agency Perspective on Electronic Submissions, 93
AIDS, 54
All-in-one software, 209–210
AMA Manual of Style, 202
Analysis of variance, 171, 175
Analytical alternatives, 183–188
Analytical measurements, 80
Anecdotal studies, 77
Angiogram review, 24–25
Animal experiments, 77, 177
Annotated abstracts, 88–91
Anti-inflammatories, 31, 33, 34, 53
Appendixes, in proposals, 82–83
Appointments, missed, 65–66
Aspirin, 6
Assignment-of-cause committee, 25
Attorney, 24
Audit trail, 117–118, 127, 139
Automated program testing, 120–121

Backups, 139–140
Baseline data, 35–36
  obtaining, 43
Baseline measurements, 16, 39, 48
Baseline values, comparison of, 179
Baycol/Lipobay, 6
Behrens-Fisher problem, 184
Bell-shaped distribution, 56, 187
Binary restenosis, 33
Binomial data, 178
Biochemical theory, 77
Blind trials, 43
Blocked randomization, 50–51
Blood filters, 6
Blood-pressure lowering agent, 33, 34
Bonferroni inequality, 190
Bootstrap, 56, 184, 188
Box and Whiskers Plot, 157, 169
Breast implants, 49–50
Budgets, 44, 105, 163–164
  hardware and software, 124

CANDA. *See* Computer-aided new drug application (CANDA)
Case controls, 62
Case-control studies, 77
Case report forms
  printout of, 117
  storing, 202
Categorical data, 166–168, 178
  comparisons of, 172–173
CDISC Submission Metadata Model, 114, 134
CDISC submission standards, 112–114
Character data, storing and retrieving, 13
Chi-square analysis, 174
Chi-square distributions, 56, 173
Cholesterol-lowering drugs, 53, 206
C-ISAM software, 213
Claims, limiting, 38–39
Client-server architecture, 132–133
Clindex software, 210
Clinical events adjudication committee, 82
Clinical experience, review of, 34–35
Clinical follow-up, 79
Clinical investigators, responsibilities of, 146
Clinical research monitors (CRMs), 20, 21, 114, 121. *See also* CRM entries
  pretrial checklist for, 146
  responsibilities of, 146, 150–152
Clinical resource centers, recruiting from, 97
Clinical significance, 193
Clinical trials. *See also* Controlled randomized clinical trials; Trials
  cost of, 62–63
  defined, 7
  exceptions to, 181–182
  guidelines for, 6
  reasons for, 5–8
  registry of, 100
Clintrial software, 214
Clintrinet software, 212
Closure, 70
CODASYL DBTG network database model, 128
Code cracking, 67–68, 162
Coding systems, 135–136
Collateral studies, 207
Committee for the Protection of Human Subjects (CPHS), 67
Committee members, primary duty of, 25
Committees, roles of, 160–161
Company responsibilities, 82
Competing events, 182
Compliance, 102
  increasing, 60
  monitoring, 103–104, 141
  staff contribution to, 103
Computer-aided new drug application (CANDA), 93
Computer-assisted data entry, 11–12, 93, 109–124
  advantages of, 109–110
  screen development in, 114–118
Computer-assisted direct data entry, 41, 63, 159
Computer-assisted product licensing applications, 93
Confidence intervals, 167–168, 186
Confidentiality, patient, 137–138
Conflicts, in committees, 160
“Contract professionals,” 95
Contract research organizations (CROs), 26, 96
Contracts, drafting, 70
Controlled randomized clinical trials, 48–52
Coronary-stenosis reducing surgical procedure, 33, 34
Corporate cost cutters, 28
Cost considerations, 62–63
Cost overruns, 151–152, 156
Costs, tracking, 14, 15
CPHS. *See* Committee for the Protection of Human Subjects (CPHS)
Critical terms, defining, 43
CRM medical monitor meetings, 147–148
CRM monitoring activities, 155
CRMs. *See* Clinical research monitors (CRMs)
CROs. *See* Contract research organizations (CROs)
Crossover design, 61
Crossovers, 181
Cutting corners, 4

Data. *See also* Missing data
  correcting, 11
  repeated tests on, 190–191
  types of, 178
  understanding, 166–170
Data analysis, 18, 27, 81, 165–194
  program suite for, 18
  software, 26–27, 214–215
Database management systems (DBMS), 132–133. *See also* DBMS Copy software
Database manager, 21, 26, 138
Database protection, 139
Databases
  multiple, 133–137
  regulatory agency access to, 93
Database server, 132

Database testing, 139–140
Data collection, 9–11
  individuals charged with, 36
Data Desk/Activ Stats/DataDesk XL software, 214
Data entry
  computer-assisted, 11–12
  training for, 122–123
Data-entry software, 22–23, 211–212, 213–214
  development of, 110–112
Data-entry technology, 137
Data in hand report, 87
Data integrity, maintaining, 118
Data management, 80, 125–140
Data management software, 16–17, 138, 212–214
Data monitoring, 70
Data security, 137–140
Data specifications table, 111
Data storage, 136
Datatrak EDC™ software, 212
Data transfer, 136–137
DB2 database, 133
DBMS Copy software, 216
DBMS. *See* Database management systems (DBMS)
Decoded results, 162
Delays, dealing with, 156
Deleterious trends, early detection of, 14
Demographics, 92
Design, alternate, 60–62
Design checklist, 43–44
Design decisions, 29–44
Design parameters, 16
Design team, 15, 19–21
Devices checklist, 141–142
Direct telephone contact method, 102
Do-it-yourself software, 211–212
Documentation, 75–94
  guidelines for, 75–76
  of software, 193–194
Documentation checklist, 142
Domain tables, 134
Doses, missed, 65
Double-blind studies, 52–54
Dow Corning, 49–50
Downsizing, 5, 28
Dropouts, 157–158, 181
Drug interactions, 38

Economic models, 63
e-CRF. *See* Electronic Case Report Form (e-CRF)
EDC. *See* Electronic data capture (EDC)
EEG review, 24–25
Efficacy measures, 15
Efficacy trials, 7
Electronic Case Report Form (e-CRF), 109, 118, 211
Electronic data capture (EDC), 5, 11, 118–119
Eligibility forms, 68, 101
Eligibility requirements, 16, 38, 60, 78, 206
Enrollment
  ethical considerations in, 101
  monitoring, 68
Enrollment report, 86–87
Equivalence, demonstrating, 58–59
Equivalence testing, 185–186
Error, most frequent sources of, 36
Erythromycin, 38
Estimates, precision of, 55–60
E-submission, 92–93
Ethical considerations, 82, 101
European guidelines, 179
Evaluation, 79
Event monitoring, 158
Exact values, collecting, 55
Examinations, follow-up, 79
Excel spreadsheet, 125–126
Exception handling, 65–71
  patient related exceptions, 65–68
Exceptions, investigator related, 68–70
Exclusion criteria, 78
Exclusion lists, 38
Executive board, 82
Expenditures, 105
  controlling, 203
  hardware and software, 124
  tracking, 44, 163–164
  variable and fixed, 203
Expenses, excessive, 151
Experimental design, 47
Exponential distribution, 56
External committees
  findings of, 159
  functions of, 160–161
External review committees, 24–26
External review panels, 25, 68

Facilities, changes in, 150
FDA regulatory requirements, 96
FDA. *See* Food and Drug Administration (FDA)
File access, 138–139
Final reports, 9–10, 75, 91–94
Fisher’s exact test, 173
Fixed costs, 203
Flat file database, 125–127
Flow chart, for software development, 112
Flow chart development process, 22
Follow-up appointments, missed, 182
Follow-up examinations, 39
Follow-up meetings, 148

Follow-up procedures, 79
Food and Drug Administration (FDA), 110
Forecasting models, 156
Formal program testing, 120–121
Formal testing team, 23
Forms
  late and incomplete, 156
  pre-defined groups in, 112
  reviewing, 149
  simple, 13
Fractional factorial design, 61
Frequency diagram, 160
Friends, recruiting, 95
F-test, 174–175
Future studies, 207–208

Gant chart development process, 22
Globalview operating system, 23
Government regulations, 12

Handheld devices, 211
Hands-on program testing, 120–121
Hardware, 23
Hardware checklist, 142
HCOL Clinical Study software, 211
Health fairs, free screening at, 101
Hierarchical database management system, 135
Hierarchical databases, 127–128
Historical databases, 62
HTML format, 88
Human trials, 177

ICH guidelines, 91
IHS Guidelines for the Format and Content of the Clinical and Statistical Sections of an Application, 179
Implementation team, 16, 21–22
Individual-to-individual variation, 47–48
Industry-sponsored trials, 96
Ineligible individuals, 181
Information recapture, 135
Informed consent, 24, 81
Informed consent form, sample, 83–84
In numero experiments, 7
Instructions, holes in, 41
Intent-to-treat regimen, 53, 70
Interest, loss of, 151
Intergel$^{\text{TM}}$, 33
Interim quality control, 159
Interim reports, 75, 86–91
Interventions
  simple, 13, 36
  variation among, 84
Introduction to the CDISC Submissions Data Domain Models, 114, 134
Investigational plan, 76
Investigator categories, 147–148
Investigator meetings, 148, 149
Investigator motivators, 152–153
Investigator panels, 23–24
Investigator-related exceptions, 68–70
Investigator relations, 109–110
Investigator responsibilities, 81–82
Investigators, recruitment efforts of, 99
Investigators checklist, 142–143
In vitro experiments, 7
In vivo experiments, 7

Journal articles, drafting and publishing, 93–94

Karnofsky Index, 178
Key fields, 134–136
Keypunch instructions, 117
Keys, 134
Kick-off meetings, 148
Kruskall-Wallace test, 175

Laboratories, paying, 14
Laboratory guidelines, 86
Laboratory results, confirmation and validation of, 159
Lead software developer, 21, 22
Lipid-lowering therapies, 38
Log-rank test, 178
LogXact software, 215
Long-term clinical trials, 30
Long-term randomized controlled clinical trials, 61
Loss adjustment, 59–60
Lung cancer data, 177

Management
  approaches to, 4
  sloppy, 5
Manuals. *See* Procedures manuals
Manufacturing specialist, 21
Marketing representative, 21
Mass recruiting, 101–102
Master checklist, 141–144
Measurements checklist, 35
Media recruitment campaigns, 100
Medical monitors, 20, 69
  role of, 145–147
Medical organizations, close liaison with, 96
Medical records, 24
Medical reviewer, 93
Medication, 42
  adjusting, 53
Meetings, pretrial, 147–148. *See also* Kick-off meetings

Menus, pull-down or pop-up, 116
MetaTrial System software, 212
Metoprolol, 33, 34
Metric data, 55–56, 166, 168, 174–176, 178, 187
Microsoft Access software, 133, 213
  data processing with, 132
Middle management, 3
Minimum relevant difference, 185–186
Missing data, 182
Monitor schedule, 149
Monitoring, for quality control, 40
Monitors, roles of, 145–147, 148–152
Morale-building meetings, 148
Motivation, physician, 97–98, 152–153
Motrin^{TM}, 31, 33
MS-SQL database, 133
Multiple databases, combining, 133–137
Multiple trials, 30
Multivariate statistical analysis, 163

Naproxin, 31
Network database model, 128
Network usage, lowering, 133
New drug applications, 93
Newsletter, 151
  for study subjects, 103
Noncompliance, 66
Noncompliant patients, 181
Nonparametric methods, 175
Normal distribution, 55–56
NPC TEST software, 215
nQuery Advisor software, 215
nQuery software, 59
Object-oriented databases, 132
Objectives, proposal, 77
O'Brien test, 184
Observer-to-observer variation, 47–48
Office-based questionnaire, 102
Open-ended reporting, 109
Oracle Clinical software, 209–210
Oracle database, 133
Oracle software, 212–213
Oral contraceptives, 38
Ordinal data, 166, 170, 178
  analysis of, 173–174
Outcome measures, 79
Outcomes, anticipated, 10
Outliers, 182

Package insert, 31
Paperless system, 3
Paradox software, 213
Parametric methods, 174–175
Participants, telephone contact with, 103–104
PASS 2000 software, 59, 215
Passive (negative) controls, 49
Patent claim, 37–38
Patient care, continued, 201–202
Patient compliance, 85
Patient confidentiality, 137–138
Patient deaths, 150
Patient follow up, 65–66
Patient ID, 134
Patient manual, 37
Patient noncompliance, 66
Patient population, stratifying, 39
Patient records, 81
Patient-related exceptions, 65–68
Patients
  excluded, 15
  informing, 102
  long-term monitoring of, 206
  losing, 59–60
  recruiting and enrolling, 17, 25, 38, 99–102
  retaining, 102–105
  run-in period for excluding, 104–105
Patients checklist, 143
Patient selection, 77–78
Patient-to-patient variation, 84
Patient withdrawals, 147
Payment, advance, 151
Pearson's chi-square statistic, 173
Permutation tests, 175
Pharmaceuticals checklist, 141–142
Pharmacologist, 21
Phase III clinical trials, 7
Physician interest, maintaining, 152–153
Physician liaison, duties of, 148–152
Physician panel, 16
Physician participation rate, 96
Physician-patient contact, 123
Physicians. *See also* Study physicians
  design-team, 19–20
  motivating, 97–98
  paying, 14
  recruiting, 95–99
  retaining, 98
Physician's procedure manual, 85–86
Physician training, 99
Placebo, 49, 52
Plan-Do-Check approach, 164
Planned closure, 40–41
Planning, importance of, 4, 100–101
Planning checklist, 71
PocketTrials software, 211
Post hoc hypotheses, 191–192
Postmarketing analysis, 206

Postmarket surveillance
  budget, review of, 202–203
Power of a test, 58
Precision
  degree of, 168–169
  estimating, 187–188
Pre-data screen development checklist, 110–112
Pre-design checklist, 29
Pre-design phase, 15
Predictors, finding, 192
Pretrial meetings, 147–148
Preventive measures, 36–37, 69–70
Primary end points, 32–35
  reporting, 180
Principal hypothesis, 31
Prior notification policy, 65
Procedures, descriptions of, 79
Procedures manuals, 37, 75, 81, 84–86
Product, lawsuit-free, 12
Product success, clinical trials and, 6
Profit considerations, 62–63
Programmers, statistical, 26
Programmer team, 22
Programming conventions, 119
Programs, stress testing, 121–122
Program testing, 119–122
Progress, monitoring, 145–153
Progress reports, 148
Project leader, 25–26
Project management software, 210
Project manager, 19
  duties of, 149
Proposal objectives, 77
Proposals, 75
  reviewing and rewriting, 42–43
Protocol
  clarity of, 91
  deviations from, 68–69, 150–151
  preparing, 75
  sample, 82
  simple, 13
  standardizing, 43
  violations of, 53, 158
Protocol-specific criteria, 38
Pull-down menus, 116

Quality assurance, 137–140
Quality control, 36–37, 80, 109, 159–160
  monitoring for, 40
Quality-of-life data, 43
Query processing, 130
Questionnaire, office-based, 102

Radio button, 114–116
Randomization
  blocked, 50–51
  stratified, 51–52
Randomized trials, 50
Reactions, severe, 150
Read-only access, 138
Recruiting efforts, targeting, 60
Recruitment, 156–159. *See also* Mass Recruiting
  care in, 96
  factors in, 99–100
  lagging, 68
  of physicians and patients, 95–105
  problems with, 100–101
Redundant variables, 61
Regulatory agencies, 12, 54
  approval of, 16
  electronic submissions to, 110
  final report to, 18
  initial submission to, 76–78
  inquiries from, 180
  submissions to, 91–92
Regulatory agency checklist, 143
Regulatory agency complications, 13
Regulatory approval, 21
  petition for, 38
Regulatory considerations, 82
Regulatory liaison, 20–21
Regulatory requirements, 16
Relational database model, 128–131
Reports, 9–11, 38
  simple, 13
  topics covered by, 165–166
"Rescue efforts," costs of, 164
Research, academic, 96–97
Resource center guide, 97
Response adaptive randomization, 62
Responsibilities, changes in, 150
Results, decoded, 162
Review committees, external, 17, 24–26
Review committees checklist, 143
Review panels, 82
  external, 25, 68
Run-in period, 104–105
Run-in strategies, 105

SaberTooth Curriculum, 189
Safety board, 82
Safety measures, 15
Safety review committee, 161
Safety trials, 7
Sample determination, formulas for, 54–55
Samples, random representative, 191–192
Sample size, 54–60
  cost and, 63
  reducing, 55
SAS Analysis, 172
SAS software, 210, 214
SAS Univariate procedure, 160

Screen development, in computer-aided data entry, 114–118
Screen-capture utilities, 23, 120–121
Secondary end points, 34–35
Secondary hypotheses, 82
Secondary response variables, 82
Security
  best approach to, 139–140
  data, 137–140
Security levels, 138–139
Selection criteria, vague, 43
Sequential tests, 162
Servers, 132–133
Short-term clinical trials, 61, 77
Side effects, anticipated, 11
Significance level, 56
Silicon implants, 5–6, 49–50
Silktest software, 215
Simpson's paradox, 186–187
Single-blind placebo run-ins, 104
Single-blind studies, 52–54
Single-site trials, 95
Site coordinators, 24, 148
Site-specific problems, 69–70
Site visits, 149
Smirnov test, 184
Software. *See also* Programmers; Programs
  choice of, 209–210, 132
  data-analysis, 214–215
  data-entry, 22–23, 110–112, 211–212, 213–214
  data-management, 212–214
  developing, 16–17, 21, 22
  permutation test, 175
  project-management, 210
  for sample size calculations, 59
  testing, 17, 22–23
  utilities, 215–216
Software checklist, 142
Software documentation, 193–194
Software selection guidelines, 111
Solutions, front-loaded, 4–5
Speech recognition software, 211
Sponsor data, 76
Spreadsheet format, 125–126
Spreadsheets, 131, 159–160
SPSS software, 214
SQL Anywhere software, 133
SQL statements, 129, 131
Staffing, 19–28
Staff turnover, 150
Standard error, 168–169
Stata software, 160, 214
Statistical analysis, 80–81, 92, 170–178
  guidelines for, 179–183
Statistical methodology, carefully chosen, 170
Statistical programmers, 26
Statistical significance, 162, 185–186, 193
Statistical techniques, 92, 174
Statistical terminology guide, 197–198
Statistical tests, 57, 180
Statisticians, 20, 25, 161
  nonagreement among, 184–185
Statistics
  erroneous, 189–192
  most favorable, 189
Statistics checklist, 189
Statistics programs, 175–176
StatXact software, 59, 171, 214–215
Storage codes, simple, 13
Stratified randomization, 51–52
Stress testing, of programs, 121–122
StudyBuilder software, 211–212
Study committees, 82
Study design, 9
  simple, 12–13
Study justification, 76–77
Study objectives, 30–32
Study participation, increasing, 100
Study physicians, 70
  recruiting, 69
Study planning, wrong, 11
Study population, 37–39, 179–180
Study protocol, table of contents of, 76
Study time lines, 13
Subgroup hypotheses, 82
Subjects, direct solicitation of, 101
Subsamples, 59
Success
  prescription for, 15–18
  staffing for, 19–28
Support, technical, 123
Surrogate response variables, 32, 33
Survival data, 178
Sybase InfoMaker software, 212
Sybase SQL Anywhere software. *See* SQL entries
SyMetric software, 214

Tables, in a relational database, 128–131
Tabular material, 92
Teaching hospitals, recruitment from, 96–97
Team roles, 27
Technical design decisions, 29
Technical support, 123
Technical writers, 21
Tertiary end points, 35

Testing
  database, 139–140
  equivalence, 185–186
  program, 119–122
Testing leads, 22–23
Testing software, 22–23
Testing summaries, 79
Test phase checklist, 143–144
Third-party facilitator, 204
Time-to-event data, 55, 176–178
Touch screen software, 211
Toxicity, investigating, 162
Training, in data entry, 122–123
Training program, 17
Treatment code
  cracking, 40, 52
  physician breaking of, 53
Treatment effects, detecting, 58
Treatment modifications, 70, 158
Treatment plan, 78
Treatment regimen, noncompliance with, 66
Treatments
  discontinuing, 70
  efficacy and safety of, 59
  failure of, 205
  successful, 204–205
Treatment sites
  number of, 60
  results of, 180
Trends, deleterious, 14
Trial closure, 201–202
Trial delays, 156
Trial design, 15–16, 47–63
Trial management, 155–164
Trial modification committee, 25
Trial results, disclosing, 202
Trial review committee, 203–204
Trials
  bringing closure to, 40–42
  conducting, 17, 25–26
  cut-off dates for, 177
  efficacy and safety of, 61
  lengthy, 152–153
  monitoring, 145–153
  preparing for, 16
  single versus multiple, 30
  termination and extension of, 161–163
  time line for, 78
Trial time line, 39
TrialWorks™ software, 210
TrialXS software, 209
t-test, 171, 174, 184
Ttest procedure, 175–176
Type I error, 56–58, 190
Type II error, 56–58
Type-and-verify field, 116–117

Uncoded data, 159
Unplanned closure, 41–42

Validity design, 61
Variability measures, 92
Variable costs, 203
Variances, differences among, 184–185
VIP patient treatment, 102–103
Volunteers, attracting, 100

Wastage, 3
Web-based data entry, 137
Westfall sequential permutation procedure, 190
Wilcoxon test, 175, 184
Winrunner software, 215
Withdrawals, 157–158, 181

Xerox, 23

Zelen statistic, 173
Zelen's test, 180