Jonathan B. VanGeest
Timothy P. Johnson
Sonia A. Alemagno *Editors*

Research Methods in the Study of Substance Abuse

Springer

Research Methods in the Study of Substance Abuse

Jonathan B. VanGeest
Timothy P. Johnson
Sonia A. Alemagno
Editors

Research Methods in the Study of Substance Abuse

Springer

Editors
Jonathan B. VanGeest
Department of Health Policy and Management, College of Public Health
Kent State University
Kent, OH
USA

Timothy P. Johnson
Survey Research Laboratory, Department of Public Administration, College of Urban Planning and Public Affairs
University of Illinois at Chicago
Chicago, IL
USA

Sonia A. Alemagno
Department of Health Policy and Management, College of Public Health
Kent State University
Kent, OH
USA

ISBN 978-3-319-55978-0 ISBN 978-3-319-55980-3 (eBook)
DOI 10.1007/978-3-319-55980-3

Library of Congress Control Number: 2017935551

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This monograph provides an introduction and overview of the most common research methods currently being employed to study substance use- and abuse-related behaviors, primarily in regards to alcohol and/or illicit drugs, with a focus on their application in advancing understanding, prevention and treatment. Substance abuse research draws both its theories and methods from a variety of other fields, and we have tried to incorporate insights from these various perspectives here. We acknowledge up front some variation in the definitions of key concepts used in the field, particularly with regard as to what constitutes substance "use," "misuse," and "abuse." Inconsistencies in terminology abound, with "misuse" and "abuse" even used interchangeably by researchers. Some even abandon theses terms all together to focus on clinical diagnostic criteria indicative of use disorders, such as those found in the latest edition of The Diagnostic and Statistical Manual of Mental Disorders. While definitional issues are addressed in the text, due to lack of consensus in the field, some discretion is given to the chapter author(s) with regard to their preference. The monograph is divided into six parts. In the first of these, two overview chapters are provided. In Chap. 1, we chronicle how research in the field has advanced over the past fifty-plus years and how multiple waves of innovation contributed to current standards and best practices. In Chap. 2, Jennifer Reingle and Timothy Akers introduce the transdisciplinary research framework known as epidemiological criminology, which is now considered a promising approach for innovation in substance abuse research.

Part II covers quantitative approaches, including randomized controlled trials in Chap. 3 (by James Swartz), sampling strategies in Chap. 4 (by Joseph Gfroerer, Arthur Hughs and Jonaki Bose), methods of primary and secondary statistical data analysis in Chap. 5 (by Adam King, Libo Li and Yih-Ing Hser), and longitudinal methods in Chap. 6 (by Brent Teasdale and Jerreed Ivanich). Qualitative and mixed methods are examined in Part III. Paul Draus presents an overview of qualitative methods in Chap. 7, followed by Henry Browstein's discussion of qualitative data analysis methods in Chap. 8. The use of geographic information systems in substance abuse research are presented by Jacqueline Curtis and Andrew Curtis in Chap. 9, and Sheryl Chatfield and Jeffrey Hallam investigate mixed methods research strategies for substance abuse research in Chap. 10.

Measurement issues are addressed in Part IV. A general overview of substance abuse assessment is provided by Timothy Grigsby, Steve Sussman, Chih-Ping Chou, and Susan Ames in Chap. 11. This is followed by Brian Perron, David Cordova, Christopher Salas-Wright and Michael Vaughn's consideration of measurement validity in Chap. 12. The use of surveys to measure substance use behaviors is reviewed by Timothy Johnson and Jonathan VanGeest in Chap. 13, and Michael Fendrich, Timothy Johnson and Jessica Becker provide an overview of the use of biological measures in Chap. 14. In Part V, challenges and special considerations in conducting substance abuse research with several subgroups of the general population are discussed. In Chap. 15, Dianne Kerr and Willie Oglesby consider issues in the conduct of adolescent substance abuse research. In Chap. 16, these same authors address substance abuse research in the LGBT Community. Sage Kim and Michael Puisis discuss the conduct of research with incarcerated populations in Chap. 17. Finally, Part IV examines policy analysis methods. John Carnevale outlines the role played by research in the formulation of substance abuse policy in Chap. 18. The economic evaluation of substance abuse and prevention programs is examined, in Chap. 19, by Willie Oglesby and Lauren Birmingham, and the general evaluation of substance abuse prevention and treatment programs is discussed by Peggy Stephens, Zili Sloboda, and Deric Kenne in Chap. 20.

As we acknowledged earlier, substance abuse research draws ideas, theories, and methods from a variety of other disciplines. As such, we found it necessary in organizing this volume to similarly reach out to experts across a variety of fields in an effort to provide a comprehensive overview of current knowledge and practices. We are sincerely grateful for the contributions of all authors whose contributions are presented in this monograph, and thank them for their patience in working through multiple drafts with us over the past several years. We are hopeful that you, the reader, will agree.

Kent, OH, USA — Jonathan B. VanGeest
Chicago, IL, USA — Timothy P. Johnson
Kent, OH, USA — Sonia A. Alemagno
November 2016

Acknowledgments

This book is made up of contributions from a talented group of academics and practitioners. Collectively, they provide insights on some of the unique challenges (and solutions) associated with the conduct of research on substance abuse. It has been a pleasure working with this group of scholars and they have our deep appreciation for their contributions to this effort. We would also like to thank our colleagues who have contributed valuable insights and recommendations to improve our final product. Specifically, we thank Peggy and Richard Stephens for their helpful edits and contributions. Additionally, we thank Carissa Smock and Matthew Nichols (Kent State University, College of Public Health) for their help in editing text. We also thank the team at Springer for their support and seemingly endless patience as we put this volume together.

Lastly, we would like to thank our spouses—Sarah VanGeest, LuEllen Doty and Tom Alemagno for their unwavering support throughout this project.

Jonathan B. VanGeest
Timothy P. Johnson
Sonia A. Alemagno

Contents

Editors and Contributors

About the Editors

Jonathan B. VanGeest, Ph.D., is a Professor in the Department of Health Policy and Management at Kent State University. His main areas of expertise include racial and ethnic disparities in health care, limited health literacy, and care outcomes associated with structural changes in medicine. He also has extensive experience in survey research methodology, program evaluation, and research on substance abuse issues among homeless persons and other marginalized risk groups.

Timothy P. Johnson, Ph.D., is the Director of the Survey Research Laboratory and a Professor in the Department of Public Administration at the University of Illinois at Chicago. His main areas of expertise include health disparities, survey methodology, and health behaviors in disadvantaged populations. Within the field of survey methodology, his work has focused primarily on sources of measurement and nonresponse error. Johnson has also invested considerable effort in examining the health behaviors of disadvantaged populations. Among the disadvantaged populations he has worked with are homeless persons, immigrants, cultural minorities, criminal justice populations, and mental health populations. Some of the health behaviors he has examined among these populations include cancer screening, nutrition, physical activity, tobacco use, alcohol consumption, and drug abuse.

Sonia A. Alemagno, Ph.D., is the Dean of the College of Public Health and a Professor in the Department of Health Policy and Management at Kent State University. She has focused her research on substance abuse and HIV/STD prevention, particularly examining public health services delivered within criminal justice settings such as prisons and detention centers. Her awards include a National Institutes of Health Career Development Award from the National Institute on Drug Abuse and the University of Akron Outstanding Researcher Award in 2005. In addition, she has been the principal investigator on several projects including research funded by the National Institutes of Health, the Centers for Disease Control, the National Institute of Justice, and the SAMHSA Center for Substance Abuse Treatment.

Contributors

Timothy A. Akers Division of Research and Economic Development, Morgan State University, Baltimore, MD, USA

Susan L. Ames School of Community and Global Health, Claremont Graduate University, Claremont, CA, USA

Jessica Becker School of Social Work, University of Connecticut, West Hartford, CT, USA

Lauren Birmingham Department of Health Policy and Management, College of Public Health, Kent State University, Kent, OH, USA

Jonaki Bose Substance Abuse and Mental Health Services Administration, Rockville, MD, USA

Henry H. Brownstein Center for Public Policy, The Wilder School for Government and Public Affairs, Virginia Commonwealth University, Richmond, VA, USA

John T. Carnevale Carnevale Associates, LLC, Gaithersburg, MD, USA

Sheryl L. Chatfield Department of Social and Behavioral Sciences, College of Public Health, Kent State University, Kent, OH, USA

Chih-Ping Chou Institute for Health Promotion and Disease Prevention Research, University of Southern California, Los Angeles, CA, USA

David Cordova School of Social Work, University of Michigan, Ann Arbor, MI, USA

Andrew Curtis GIS Health & Hazards Laboratory, Department of Geography, Kent State University, Kent, OH, USA

Jacqueline W. Curtis GIS Health & Hazards Laboratory, Department of Geography, Kent State University, Kent, OH, USA

Paul Draus Department of Behavioral Sciences, The University of Michigan-Dearborn, Dearborn, MI, USA

Michael Fendrich School of Social Work, University of Connecticut, West Hartford, CT, USA

Joseph Gfroerer Substance Abuse and Mental Health Services Administration, Rockville, MD, USA

Timothy J. Grigsby Department of Kinesiology, Health, and Nutrition, University of Texas, San Antonio, TX, USA

Jeffrey S. Hallam Department of Social and Behavioral Sciences, College of Public Health, Kent State University, Kent, OH, USA

Yih-Ing Hser University of California, Los Angeles, CA, USA

Arthur Hughes Substance Abuse and Mental Health Services Administration, Rockville, MD, USA

Jerreed Ivanich Department of Sociology, University of Nebraska-Lincoln, Lincoln, NE, USA

Deric Kenne Department of Health Policy and Management, College of Public Health, Kent State University, Kent, OH, USA

Dianne L. Kerr College of Education, Health, and Human Services, Kent State University, Kent, OH, USA

Sage Kim School of Public Health, University of Illinois at Chicago, Chicago, IL, USA

Adam King University of California, Los Angeles, CA, USADepartment of Mathematics and Statistics, California State Polytechnic University, Pomona, CA, USA

Libo Li University of California, Los Angeles, CA, USA

Willie H. Oglesby College of Population Health, Thomas Jefferson University, Philadelphia, PA, USA

Brian E. Perron School of Social Work, University of Michigan, Ann Arbor, MI, USA

Michael Puisis Correctional Consultant, Chicago, IL, USA

Jennifer M. Reingle Gonzalez Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas School of Public Health, Dallas, TX, USA

Christopher Salas-Wright School of Social Work, The University of Texas at Austin, Austin, TX, USA

Zili Sloboda Applied Prevention Science International, Inc., Ontario, OH, USA

Peggy Stephens Department of Social and Behavioral Sciences, College of Public Health, Kent State University, Kent, OH, USA

Steve Sussman Institute for Health Promotion and Disease Prevention Research, University of Southern California, Los Angeles, CA, USA

James A. Swartz Jane Addams College of Social Work, University of Illinois at Chicago, Chicago, IL, USA

Brent Teasdale Department of Criminal Justice and Criminology, Georgia State University, Atlanta, GA, USA

Michael G. Vaughn School of Social Work, Saint Louis University, Saint Louis, MO, USA

Part I

Overview

History of Substance Abuse Research in the United States

1

Jonathan B. VanGeest, Timothy P. Johnson, and Sonia A. Alemagno

1.1 Introduction

Substance abuse is a global problem of epidemic proportions (Degenhardt and Hall 2015; Gowing et al. 2015; Griffiths et al. 2008). In the United States alone, an estimated 17 million people are dependent upon or have abused alcohol in the past year, with males at greatest risk, as well as young adults aged 18–25 years compared to other age groups (CBHSQ 2015). An estimated 7.1 million people aged 12 or older are dependent on or abused illicit drugs in the past year, with rates of abuse highest for males, young adults (18–25 years of age), and African Americans (CBHSQ 2015). While rates have been somewhat stable over the past five years, some analysts indicate that in 2014, past month illicit drug use and alcohol dependence were at the highest rate in more than a decade, with the increase driven primarily by escalations in marijuana use, nonprescription drug abuse, and heroin use among adults 25 years of age and older (ASAM 2015).

Prescription drug abuse, in particular, has exploded in the U.S. over the past two decades, especially among older teenagers and young adults (Dart et al. 2015; Lankenau et al. 2012; Maxwell 2011; Paulozzi et al. 2011; Sung et al. 2005; Young et al. 2012). Recent increases in heroin use and the emergence of new synthetic drugs—some of which are 10,000 times more powerful than morphine—have also been problematic (Abbott and Smith 2015; Palamar et al. 2015; Palamar and Acosta 2015). With regard to tobacco use, current national prevalence rates hover around 25%, despite declines in cigarette smoking among adults, with use varying by geography and sociodemographic factors, such as gender, age, race, ethnicity, and socioeconomic status (Agaku et al. 2014; King et al. 2012). Progress has also slowed in recent years due in part to the expanded use of alternative tobacco products, such as smokeless tobacco, cigars, hookah, and e-cigarettes, especially among youth and young adults (Agaku et al. 2014; Lee et al. 2014; McMillen et al. 2012; Singh et al. 2016).

The individual and public health implications of substance abuse are significant, with adverse consequences including, but not limited to, overdose death, education and vocation impairment, accidents, violence, developmental harms to children, and increased rates of a number of diseases, including HIV infection, heart disease, cancer, and tuberculosis (Paulozzi et al. 2011). In many instances, those suffering either directly or

J.B. VanGeest (✉) · S.A. Alemagno
Department of Health Policy and Management, College of Public Health, Kent State University, Kent, OH 44240, USA
e-mail: jvangees@kent.edu

S.A. Alemagno
e-mail: salemagno@kent.edu

T.P. Johnson
Survey Research Laboratory, College of Urban Planning and Public Affairs, University of Illinois at Chicago, Chicago, IL 60607, USA
e-mail: timj@uic.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_1

indirectly from the consequences of substance abuse are at additional risk themselves of developing further addictions to alcohol and/or drugs. In addition to the aforementioned health implications, substance abuse also represents a significant economic burden to society. In the U.S., the estimated economic costs associated with heavy alcohol consumption alone are in excess of $200 billion annually (Bouchery et al. 2011; NIDA 2016). Tobacco and drug addiction are also costly, with overall annual costs estimated to exceed $480 billion (NIDA 2016). The estimated total cost of just one emergent problem—the nonmedical use of prescription opioids—amounts to $50 billion annually (Hansen et al. 2011).

A snapshot of the substance abuse issue globally and in the U.S. would be impossible without ongoing research and public health surveillance. Worldwide, a variety of primary data sources inform substance abuse and dependence prevalence rates, respectively, with considerable variation in quality and coverage of available data across countries (Degenhardt et al. 2011, 2014; Gowing et al. 2015). In the U.S., large-scale representative surveys, such as the household-based National Survey on Drug Use and Health and the school-based Monitoring the Future (MTF) studies, provide rich data on substance abuse in the United States; pointing to ongoing shifts in trends of abuse critical for an informed and effective public health response. Additional methodological approaches, primarily social surveys and epidemiological studies, further inform our understanding of substance abuse, the nature and mechanisms of addiction, and an appropriate solution. The appropriateness of this response is largely dependent upon the quality of data provided, which is in turn based on the soundness of the methodological approaches employed in its collection.

At various times, substance abuse research has produced conflicting conclusions, due in part to inherent limitations in the methods employed (Turner and Miller 1997; Westermeyer 1990). Thus, the historical context of addiction research is critical to our understanding of how to address this problem. While not necessarily linear, advancements in substance abuse research methodology build upon one another, expanding our understanding of addiction, while informing treatment and policy solutions. It has even been argued that methodological advances in this one area of inquiry—research on tobacco particularly—have advanced and shaped the field of epidemiology generally (Samet 2016). Throughout this chapter, we explore the history of substance abuse research in the U.S. while noting some promising recent developments in the methods employed.

1.2 Definition of Key Terms

Various definitions with regard to what constitutes substance "use," "abuse," and "misuse" are common in the research literature. This variation must be addressed in any discussion of the history of research in this area, as these inconsistencies have muddied the waters some in terms of research on alcohol and drug-related problems. For instance, "use" typically refers to the use of drugs or alcohol without consideration or assessment of frequency of use, risk of dependence and/or health-related consequence. Substance "abuse," on the other hand, generally refers to the use of any substance that may be unlawful and/or detrimental to the user. Frequency of use (e.g., regular or persistent use) or user intent to obtain psychotropic effects are also considerations in some definitions of abuse. Lastly, "misuse" is generally defined by emphasizing the use of substances in a manner contrary to medical indication or prescription (Smith et al. 2013). The concept of misuse gained traction, particularly in the 1980s, following concern by some that the term "substance abuse" represented an overly pejorative label and with the consideration that abusive action was not actually perpetuated on the drugs themselves (substances are

used or *misused* in terms of time- and place-bound values, but living organisms can be *abused*).[1] Over time, however, the concepts of abuse and misuse have often been used synonymously by researchers, as both involve use that contradicts medical advice (Smith et al. 2013). Choice of terminology is not simply an issue of semantics, as there are potential measurement implications associated with the construction utilized. As a result, researchers have periodically sought to better clarify and/or standardize the terminology employed in the field. In the mid 1980s, a four-stage Delphi survey of experts was conducted seeking to gain greater clarity and uniformity in terminology associated with drug/alcohol-related problems (Rinaldi et al. 1988). Achieving distinctions between misuse and abuse has been particularly important for researchers examining nontherapeutic use of prescription drugs, with the Analgesic, Anesthetic, and Addition Clinical Trials, Translations, Innovations, Opportunities, and Networks (ACTTION) public-private partnership convening an expert panel to develop mutually exclusive classifications (Smith et al. 2013). Despite efforts, definitional tensions and inconsistencies have persisted. For the purposes of this chapter, we follow the more general convention of referring to substance abuse, with additional reference to misuse, particularly related to research on prescription drugs.

1.3 History of Substance Abuse Research in the United States

1.3.1 Early Addiction Research

Due to the interdisciplinary nature of substance abuse research, establishing an accurate historical timeline for the field's development is somewhat difficult. However, there are some clear benchmarks that can be reasonably ascertained. For instance, most scholars would agree that addiction research was in its infancy during the late nineteenth and early twentieth century. Moreover, research activity during this period was sporadic, initially focused on understanding and addressing the observed associations between substance abuse and crime, thereby informing the progressive era reforms that aimed to rid society of the social problems associated with the consumption and abuse of alcohol and drugs (Campbell 2007). Research was typically conducted in private clinical settings, and was largely uncoordinated in any meaningful fashion. Often taking cues from the work of Sigmund Freud and others, early scientific investigations probed as to whether there were physiological and/or psychological markers capable of differentiating addicts from non-addicts (Acker 2002; Campbell 2007).

Utilizing a variety of research methodologies, ranging from laboratory experiments involving humans and animals to surveys and clinical chart reviews, scientists and practitioners sought to better understand the mechanisms of addiction. While advanced for their time, not all of the methodologies employed would be considered scientifically rigorous by today's standards. The work was also oriented by multiple and often competing explanatory models of addiction, many of which viewed the condition somewhere on a continuum between vice (a moral issue, with addiction resulting from emotional or psychological defects) and disease (the medical model of abuse and addiction) (Campbell 2007). In the U.S., much of the early research was funded by the Public Health Service (PHS) and the Rockefeller Institute's Bureau of Social Hygiene. The latter was established in 1913 principally as a grant-making agency emphasizing research and education. The Bureau was charged with "the study, amelioration, and prevention of those social conditions, crimes, and diseases which adversely affect the well-being of society" (Rockefeller Archive Center). While both alcohol and drug addiction fell within this charge, the Bureau specifically focused on narcotics, a choice driven largely by changing social attitudes toward the substantial rise in opium consumption (and later morphine, heroin, and cocaine) for

[1]Personal communication with Stanley Einstein, Founding Editor (1965–2013) of *Substance Use & Misuse*, November 2016.

medicinal and nonmedicinal purposes (Acker 1995; Musto 1997). Concurrently, mainstream alcohol and alcoholism studies, despite having a relatively strong scientific focus in the latter part of the nineteenth and the early twentieth centuries, were virtually halted following the passage of prohibition (Roizen 2000).

In 1921, the Bureau of Social Hygiene established the Committee on Drug Addictions, in order to stimulate research for new drugs possessing reduced addiction liabilities that might be substituted for opium-derived analgesics in medical practices. The Committee expanded their work, initially agreeing on a three-pronged effort that combined educational efforts for clinicians with sociological and laboratory research (May and Jacobson 1989). While the latter was dedicated to the search for alternate drugs, which ultimately became the focus of much of the committee's work, the sociological research sought to better understand the extent of the problem, as well as mechanisms of drug distribution and the economic implications of addiction, often utilizing surveys (Acker 1995, 2002). Shifting government policy towards a view of addiction from a criminological standpoint increasingly stifled research on addiction from a more medical/disease orientation for fear that it would undermine government policy and the criminal justice solutions being enacted (Acker 2002; Musto 1996, 1997). Both the PHS and Bureau of Social Hygiene supported this shift, with leadership increasingly framing addiction as a problem of criminology and/or vice, consistent with federal policy (Acker 2002). As such, the subsequent orientation clearly located the etiology of addiction in an individual's psychopathology (Acker 1997).

Research funding by the Bureau and the PHS also helped galvanize consensus on the status of addiction research through the formation of networks of scientists oriented toward research that might best improve the understanding and control of social problems associated with the abuse of alcohol and other drugs (Acker 2002). Although academic in nature, this research had very practical applications consistent with the focus on social reforms that emerged during this period (Acker 2002). Despite advancements in funding and corresponding sophistication in the research methodologies employed, including new city-wide surveys and advanced physiological studies, overall consistency in the quality of methods employed in the field as a whole remained elusive.

While researchers were active, overall scholarly productivity remained relatively low until 1929, when research on drug addiction was first mandated by the United States Congress (Campbell 2007; Musto 1996, 1997). That same year, the Committee on Drug Addiction was established by the National Research Council (NRC), in order to provide a more strategic and systematic approach to addiction research, including plans for a coordinated program of chemical, pharmacological, and clinical research. The Bureau of Social Hygiene also decided to transfer its support of research to the NRC, furthering this strategic reorganization. The NRC initially sought to identify key gaps in biological knowledge regarding addiction, and while both alcohol and other drugs were included under this broad charge, the Committee initially focused on morphine, in part to continue the search for an alternative to the drug that was not habit forming (Acker 1995; Musto 1996). Despite the passage of the Harrison Narcotics Act in 1914, which restricted morphine use, and the banning of all domestic manufacture of heroin in 1924, the drug remained one of the most commonly abused narcotics. Studies were conducted in both laboratories and clinical settings involving human subjects, initially at Yale and the University of Virginia. Animal studies were also initiated in a newly developed pharmacology unit at the University of Michigan (Acker 2002; Musto 1996).

Additional clinical facilities were set up in Lexington, Kentucky and Ft. Worth, Texas in the mid- to late-1930s. Christened "narcotic farms," they were actually special prisons for drug addicts, which were maintained under the supervision of the Public Health Service (Acker 1997; Musto 1996). At these prisons, clinical studies were conducted on prisoners to examine the compounds developed and produced in the new laboratories. This shift from a three-pronged

strategy to exclusively focusing on clinical and laboratory studies was done largely at the expense of sociological studies, which the Committee did not view as particularly helpful in addressing the drug situation (Musto 1996). However, social science research continued during this period albeit outside of the Committee, focusing on understanding the behavior, as well as the social determinants, that influence the abuse of drugs and alcohol.

Ultimately, the work of the NRC Committee on Drug Addiction gave birth to the Addiction Research Center (ARC) located in Lexington, Kentucky (Musto 1996). The ARC was established in 1935 as one of the first research laboratories in the National Institute of Mental Health. At the time of its establishment, the ARC was the only laboratory in the world devoted solely to the study of addiction, and one of the earliest explicitly multidisciplinary laboratories in existence (Campbell 2010). While the ARC was primarily charged with the study of the clinical neurophysiology of drug dependence, it also produced innovative research methods to examine potential public health problems associated with addiction (Campbell 2006, 2010; National Institute on Drug Abuse 1995). Still affiliated with a Federal prison in Lexington, the ARC provided treatment and conducted research on prisoners and others who voluntarily admitted themselves to the facility. Among the many successes attributed to the ARC is furthering the understanding of relapse behavior and the profiling of the physiological and psychological effects of different drug classes. Pharmacological research also provided major contributions to the understanding of opiate and alcohol dependence and withdrawal, as well as expanded opportunities for the advancement of new drug development, with new research methods devised in partnership with industry and academic partners to test the pharmacological effects of novel compounds (Acker 1995).

Corresponding abuse liability studies were developed, again utilizing innovative experimental methods, to assist scientists in determining whether new pharmaceutical products were addictive, or whether they might have the potential for treating addiction and abuse (Acker 1995). Use of prisoners allowed for the development of closely comparable research protocols to test for addictiveness. All of this work coalesced around a model of addiction stressing the psychoneurotic individual possessing preexisting defects of personality that predisposed them to intractable addiction, emerging as the dominant explanatory model of the period. The continued linkage between addiction research and social reform, as noted earlier, reintroduced opportunities to pursue disciplinary agendas beyond exclusively pharmacological approaches, including new and innovative multidisciplinary addiction research incorporating biomedical sciences, social sciences, and public health (Campbell 2010).

Given the expansion of focus and resources, research on alcohol and tobacco use developed apace with drug research during the 1920s and 1930s. Studies on alcohol and alcoholism commenced again following the repeal of prohibition in 1933 (Roizen 2000), with early research focusing primarily on the metabolism and physiology of alcohol consumption (Candon et al. 2014). In 1938, the American Association for the Advancement of Science founded the Research Council on the Problems of Alcohol, an association of scientists and doctors whose goal was to raise support for multidisciplinary research on the effects of alcohol on the body, in addition to studying the extent of alcoholism in U.S. society. In 1939, the Council hired E.M. Jellinek to conduct the first comprehensive review of the literature on the effects of alcohol on the individual (Candon et al. 2014; Weglarz 1987).[2]

Dissemination of research findings was also a key component of the Research Council's activities. Their official journal, the *Quarterly Journal of Studies on Alcohol*, was founded in 1940 by Howard W. Haggard, M.D., director of Yale University's Laboratory of Applied Physiology, as the only scientific periodical at the time

[2]Initial funding for the review was from a grant from the Carnegie Corporation. Designated as the Classified Abstract Archive of Alcohol Literature (CAAAL), the collection was maintained and updated until 1978 and is comprised of approximately 20,000 abstracts.

devoted solely to the study of alcohol and alcoholism. The Laboratory was already heavily involved in alcohol research when, in 1943, it established the Section of Studies on Alcohol to further expand its body of work, made up of a multidisciplinary research team of sociologists, psychologists, physicians, biochemists, and economists (Candon et al. 2014; Page 1988, 1997; Roizen 2000). Research methods were a major focus of this new section, with statistical methods explored to produce improved, and oftentimes controversial, measures of the prevalence of alcohol-related phenomena, including measures of use and new alcoholic typologies (Page 1997). This work was also instrumental in achieving the recognition of alcoholism as a major public health problem, as well as a treatable illness, in the face of open social hostility toward both the alcoholic and the addict (Page 1997; Roizen 2000; Warren and Hewitt 2010).[3]

On the tobacco front, early epidemiologic studies linking tobacco use and cancer were occurring largely outside of the United States (Cummings 2002; Proctor 2012; Samet 2016). By the 1930s, experimental studies in South America and Europe had led many researchers and clinicians to conclude that smoking was indeed a potential cause for a number of cancers (Doll 1998; Proctor 2012). These initial studies were often case-control designs, implemented in cooperation with clinics and hospitals. Tobacco use was typically ascertained using questionnaires, and while advanced compared to previous designs, many of these studies had methodological limitations (Doll 1998; Proctor 2012; Samet 2016). These limitations led many in the U.S. to ignore or dismiss the results. Also important was public opinion, which was shifting as most state and local prohibitions against tobacco use were being lifted in favor of taxation policies, opening up opportunities to market tobacco products to wider audiences and ultimately contributing to a period of unprecedented growth in the prevalence of smoking (U.S. Centers for Disease Control and Prevention 1999, Cummings 2002).

Despite clear methodological and scientific advances in addiction research in the 1920s and 1930s, a non-habit-forming analgesic had still not been found. While many drugs had been tested, the methodologies employed were still quite simple by today's standards, often involving merely substituting the test drug for a regular dose of morphine in a morphine-addicted person, with subsequent observations failing to address the molecular level where dependence actually occurs (Musto 1996). Without the driving need to identify a non-habit forming alternative for medical purposes, alcohol research languished alongside treatment and support programs, which remained underfunded and underdeveloped during this period (Warren and Hewitt 2010). Tobacco use was even encouraged by the government, which was including cigarettes in the rations for soldiers during World War II. However, a foundation for future addiction research had been laid, and notable advances in substance abuse research achieved. Methods had been improved as researchers began to move toward more sophisticated designs; documenting the pitfalls of drug testing and, by all accounts, making significant progress in advancing addiction research methods along multidisciplinary lines. While addiction research was still in its infancy, the progress was striking, especially considering its rather modest beginnings as a more coherent science just a few decades previous.

1.3.2 Post War to 1965—A New Beginning in Substance Abuse Research

Addiction research during the war years was largely on a hiatus, although some clinical studies supporting new drug development continued in Lexington. After the war, the Committee on Drug Addiction and Narcotics (CDAN) was established by the NRC in 1947 to replace the Committee on Drug Addiction. Research again focused on drug development, with initial studies

[3]This shift recognizing alcoholism as a treatable medical condition occurred despite the corresponding loss of faith that narcotics addicts could be similarly treated; a view that persisted into the 1960s, with implications for research.

concentrating on methadone, a synthetic analgesic developed by German scientists (May and Jacobson 1989). Researchers' considerable interest in methadone's possibilities prompted requests to pharmaceutical manufacturers to contribute to a designated research fund that the Committee would administer (Musto 1996). University science departments also contributed some of their own resources, along with other outside agencies, including the Veterans Administration and the World Health Organization. This fund grew quickly, allowing for the sponsoring of a variety of research, including studies of methadone as well as other synthetic opioids and opiate antagonists, the latter referring to drugs that block opioids by attaching to the opioid receptors without activating them (May and Jacobson 1989).[4]

Still, no analytical techniques were developed that were sufficiently sensitive or specific to measure levels of opiates and/or similar compounds in blood or urine, forcing researchers to rely primarily on clinical observation. However, within these limitations, advanced research methods were employed, including double-blinded techniques, to compare the effects of new drugs with those of a placebo and the standard drug, which was often morphine, ultimately serving as models for future clinical drug trials. The pharmacological research at the Lexington facility provided major contributions to the understanding of opiate and alcohol dependence and withdrawal, building upon what amounted to decades of baseline data by the early 1950s (Campbell 2007, 2010; May and Jacobson 1989). Much of the Lexington research was still conducted using prisoners during this period, although pending legislation was poised to fundamentally alter the relationship between the larger prison-hospital and the research unit (Campbell 2010). Aside from research, the Committee served in an advisory role to agencies, such as the Federal Bureau of Narcotics and the Food and Drug Administration (FDA), informing on the potential abuse liability of marketable drugs.

In addition to the work of the CDAN, a number of other factors contributed to a resurgence of post-war addiction research. In 1949, the National Institute of Mental Health (NIMH) was established as one of the National Institutes of Health, providing additional coordination and funding opportunities for substance abuse research.[5] By the early 1960s, private foundations had also begun to fund addiction research, prompted by an increasing concern over the rise of illicit drug use. This concern also prompted Federal action, including the 1962 White House Conference on Drug Abuse and the subsequent report of the President's Commission on Narcotics and Drug Abuse released the following year (Musto 1996). The result was an expansion of work far beyond the more narrow focus on drug development that characterized much of the research in previous decades.

Expanding illicit drug use also spurred more localized research initiatives, especially in large urban centers where community-level solutions were sought to counter the emergent drug culture. This included social (community-based) research, such as ethnographic studies which sought to understand drug addiction within the context of culture, as well as pharmacological and clinical examination (Neale et al. 2005). For instance, in New York, concern over heroin addiction prompted Rockefeller University to partner with the New York City Health Research Council to conduct pharmacological research, with the goal of developing appropriate classifications for

[4]For a discussion of the various drugs tested by the CDAN, as well as Committee composition, see the detailed narrative history of the Committee on Problems of Drug Dependence by May and Jacobson (1989).

[5]The new agency adopted a model approach to mental disorders, including addiction, which stressed the interrelatedness of research, training, and services. As a result, the research portfolio of the NIMH differed significantly from other NIH institutes. In addition to basic and clinical biomedical research, NIMH strongly supported behavioral research and some social science research. The three-pronged approach, however, did create inherent tension, as the combination of research and service in a single agency left advocates for each side concerned that they may not be receiving equal prioritization of funding and support. This tension would remain until being resolved in later decades.

addicts and improved options for managing the problem (Kreek et al. 2004). These initiatives helped refine the methods of inquiry; both in terms of study designs employed, and also in the community-based translational research projects that followed as scientists sought an effective pharmacotherapy for heroin addiction that could be combined with behavioral care (Kreek et al. 2004). Central in this expansion and search for novel treatments was an emerging shift in orientation toward narcotic addiction. As noted previously, most scientists at this time still considered drug addiction as either deviant behavior or the result of a personality disorder (Campbell 2007). However, community-based researchers, experienced in working with addicts in New York and elsewhere, realized that arrest and incarceration were not effective methods of management, and began to thus reframe narcotic addiction once again as a disease.

Alcohol research also accelerated in the 1950s and early 1960s, with new methods employed to better understand abuse. For example, an early longitudinal study conducted between 1949 and 1952 collected data on 16,000 college students from 27 colleges and universities in the United States explored the etiology of addiction within this population (Fillmore and Marden 1977). Prospective and longitudinal study designs were also improved and used to examine the stages and patterns of alcohol abuse, with attention to potential antecedents of change signaling transitions from non-problematic to problematic drinking behavior (Burgheim 1953; Fillmore and Marden 1977). Qualitative inquiry, in particular, was resurgent in this era, as sociologists, social psychologists, and anthropologists sought to better understand addiction in the context of culture, bringing with them various ethnographic methods and approaches to studying alcohol abuse (mirroring what was happening in drug addiction research; see also Chap. 6 in this volume).

In the 1950s, the Yale Laboratory of Applied Physiology's Section of Studies on Alcohol was rebranded as the Center of Alcohol Studies, the first multidisciplinary research institution focusing explicitly on alcohol problems (Campbell 2007; Page 1997). Initial consideration was given primarily to the sociological aspects of abuse, with the physiological and psychological aspects the purview of a new section within the Laboratory of Applied Physiology, the Laboratory of Applied Physiology, and Biodynamics (Candon et al. 2014); in 1961, the Center moved to Rutgers University. At the same time, there was an emergent view of alcohol dependence as a separately recognized medical disorder, reinforcing the "disease concept of alcoholism" and shaping future research (Jellinek 1960). Alcoholism research itself was also emerging as a legitimate science, with a move afoot to create an institute within NIH solely dedicated to the alcohol field (Israel and Lieber 2002; Page 1997). Still lacking urgency, progress was still slow by the mid-1960s, with the NIMH beginning a small grants program in the area of alcohol research, and establishing the Center for Prevention and Control of Alcohol Problems, though these initiatives had limited budgets and/or authority (Warren and Hewitt 2010). Overall, stable support for alcohol research was summarily lacking, with funding for studies often cobbled together from a variety of sources, including government agencies, charitable organizations, and industry, slowing the overall progress as a coherent field of inquiry (Candon et al. 2014; Warren and Hewitt 2010).

Major breakthroughs in tobacco research did occur in the early 1950s and 1960s when scientists from the United States and elsewhere began to publish their research linking smoking and cancer, thereby birthing the modern era of tobacco control (Doll 1998; Parascandola 2001; Proctor 2012; Samet 2016). By the end of 1953, thirteen epidemiologic studies linking smoking to cancer had been completed, most of which utilizing case-control methods (Parascandola 2004). Advances in methodological rigor were evident in this work, including the development of new statistics to assess risk. Results sparked critical debate among researchers, with some concluding that there was sufficient evidence to conclude a cause and effect relationship, while others remained skeptical in the face of available evidence. In a review of this debate,

Parascandola (2004) notes that methodological weaknesses of the case-control method played an important role in this dialogue.[6]

Parascandola further mentions that improved study methods, including the first large-scale prospective cohort studies, ultimately strengthened the body of evidence against tobacco use, but that the debate continued, as these new methods also had weaknesses, including selection bias and an inability to control for potential hidden confounders (Parascandola 2004). Anomalies in the mass of evidence, such as the lack of association between cigar and pipe smoking and cancer, were likewise problematic. While not fully resolved among all investigators, the growing tide of evidence, including new randomized controlled trials, began to overwhelm the argument. In a 1964 groundbreaking report, the Surgeon General's Advisory Committee on Smoking and Health concluded, based on the consistency, strength, and coherence of the available evidence, that "Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction" (Advisory Committee to the Surgeon General of the Public Health Service 1964).

Despite being an era of new beginnings for substance abuse research, the period between 1950 and the mid-1960s has been characterized by some as the infancy of substance abuse research (Campbell 2007). Specifically, the continued prioritization of criminalization over treatment, especially with regard to drug addiction, hindered individual and private research. Other factors also played significant roles, including computational and methodological study limitations, resulting in the continued reliance on largely observational study methods, as well as the failure of a stable funding base to emerge from which to launch addiction research on a larger scale. Moreover, most studies were only conducted on men. Lastly, studies were hampered by an over-reliance on unitary models of dependence that failed to fully articulate the mechanisms of addiction (Nathan and Lansky 1978). Where significant advances were noted, such as in the establishment of the tobacco-cancer link, they were largely foundational, setting the stage for major future advances.

1.3.3 1965—Today

Funding for drug abuse research expanded dramatically in the 1960s and early 1970s, due to increases in grants by NIMH. Also facilitating this increase were evolving public attitudes around drug addiction, which began to support treatment research with individuals struggling with addiction as opposed to punishment (Musto 1996). The Committee on Drug Addiction and Narcotics changed its name to the Committee on Problems of Drug Dependence (CPDD) in 1965 to meet the new definition of "addiction" promulgated by World Health Organization and others, which explicitly viewed illicit drug abuse as a disease (May and Jacobson 1989; Musto 1996).[7] In 1967, the Center for Studies of Narcotics and Drug Abuse was formed within NIMH to administer the rapidly growing portfolio of grants and contracts dedicated to the study of problems related to drug abuse. A year later, NIMH's new Division of Narcotic Addiction and Drug Abuse assumed administrative responsibilities for all of the agency's research activities. Further abandonment of the punitive-deterrent

[6]Also contributing were favorable research studies funded by tobacco companies and/or comments by scientific experts discounting the evidence that were part of a broader marketing and public relations campaign designed to challenge evidence that smoking caused disease (Cummings et al. 2007).

[7]The WHO's new definitions facilitated their increased responsibility, as established by international treaties, to control narcotics. In the 1950s, the presence of physical dependence was emphasized, with the WHO primarily concerned with differentiating between psychic dependence and physical dependence. In 1969, the WHO abandoned efforts to differentiate habits from addictions and adopted terminology designating as dependence "those syndromes in which drugs come to control behavior." They further recognized that dependencies on different classes of drugs (such as alcohol, opiates, cocaine) can differ significantly.

philosophy in the U.S. followed the report of the President's Commission on Narcotics and Drug Abuse, which advocated adoption of approaches in line with the view of illicit drug abuse as a disease. Congress followed by enacting the Comprehensive Drug Abuse and Control Act of 1970, establishing the National Commission on Marijuana and Drug Abuse, which would report on a range of issues linked to drug use, arguably the most important (from a researcher's perspective) being the Commission's second report, as it promulgated strong recommendations for the expansion of government-sponsored research and for the continuance of newly implemented national surveillance surveys on drug use, including the National Household Survey on Drug Abuse (Musto 1996).[8]

The Commission further conceived a wider range of research relevant to drug issues to be incorporated into the research programs of the NIMH. This research, along with all of the NIMH intra- and extramural treatment and research activities was transferred to the National Institute on Drug Abuse (NIDA) following its formation in 1974 (Kreek et al. 2004; Musto 1996).[9] Today, NIDA supports most of the world's research on the health aspects of drug abuse and addiction, with strategic research and treatment priorities focused on priority areas that include better understanding of the factors influencing drug use trajectories, accelerating the developments of new treatments, and supporting translational research to ultimately improve individual and public health (NIDA 2015; Sloboda 2012).

While NIDA is the dominant funder, other Federal agencies, such as the U.S. Centers for Disease Control and Prevention and the U.S. Department of Justice, also fund drug research, with the latter focusing on drug use in relation to violence and crime (Sloboda 2012).[10] In addition to government funding, foundation support for drug abuse research and treatment also emerged in the 1960s and 1970s. An early example is The Ford Foundation's Drug Abuse Survey Project, which sought to identify gaps in basic knowledge of drug addiction and the role of drugs in society, resulting in the Foundation's creation of a Drug Abuse Council, which funded studies on illicit drug abuse from 1972 to 1978 (Musto 1996). Although relatively few foundations focus exclusively on substance abuse research, compared to its modest beginnings, foundation backing of drug abuse research overall helped to stabilize research support, as well as promote the integration of evidence into treatment in subsequent decades as new drug-related issues emerged, such as the Crack epidemic of the 1980s or today's extra-medical use of prescription drugs (Acker 2002; Musto 1996; Sloboda 2012).

Acknowledging the need for more information on problem drinking, President Nixon signed the Comprehensive Alcohol Abuse and Alcoholism Prevention, Treatment, and Rehabilitation Act in 1970, authorizing a comprehensive Federal program to address prevention and treatment of alcohol abuse and alcoholism, including the expansion of alcohol addiction research (Warren and Hewitt 2010). Alcoholism was also acknowledged as a serious, but curable, public health problem. Moreover, the National Institute on Alcohol Abuse and Alcoholism (NIAAA) was established as a component of the NIMH; subsequently becoming a separate institute alongside NIDA in a move that further increased targeted funding for alcohol research

[8]The National Commission on Marijuana and Drug Abuse's second report is entitled Drug Use in America: Problem in Perspective (NCMDA 1973).

[9]The creation of NIDA in 1974 was a major step forward in the promotion of addiction research, as previous work had been folded into the larger portfolio of the National Institute on Mental Health (Kreek et al. 2004; Sloboda 2012). The new institute focused exclusively on drug research. In 1992, NIDA became part of the National Institutes of Health.

[10]NIDA also funds drug and crime research. Examples include the NIDA funded research conducted by the National Development and Research Institutes examining the relationship between drugs and criminality (Lipton and Johnson 1998). Today, NIDA is increasingly focused on medical interventions and brain science research.

(Candon et al. 2014; Israel and Lieber 2002; Warren and Hewitt 2010).[11]

As the lead Federal agency addressing problems associated with alcohol abuse and alcoholism, NIAAA primarily supported research to improve understanding of the scope and nature of alcohol addiction and its effects on the body, as well as exploration of new alcoholism treatments (Lieber 1989; NIAAA 2011; Warren and Hewitt 2010; Willenbring 2010). The agency also backed efforts to prevent alcohol-related problems through policy research and scientific support for advocacy, including targeted efforts addressing underage drinking, college drinking, and parental alcohol exposure, among others (Voas and Fell 2010; NIAAA 2011). What followed was a proliferation of new research and research centers across the country, such as the Research Society on Alcoholism, thus furthering the organization and expansion of alcohol and alcoholism research (Israel and Lieber 2002; Lieber 1989).

Tobacco research also grew in recent decades. However, unlike other areas where one or (at most) two agencies were primarily responsible for spearheading the federal response, numerous agencies promoted research on nicotine addiction and tobacco use, including the National Cancer Institute (NCI), the FDA, the U.S. Centers for Disease Control and Prevention (CDC), the Office of the Surgeon General, NIDA, and the Agency for Healthcare Research and Quality (AHRQ), each with different focuses and priorities.[12] Much of this work has focused on better understanding general patterns and determinants of use, as well as developing more comprehensive epidemiologic models for understanding tobacco addiction and its impact on health (Doll 1998; Giovino 2002). Additional topics included tobacco use by women, adolescents, and other minority and underserved populations, better understanding patterns of addiction and related health risks, and addressing key deficiencies in the knowledge base noted in earlier reviews and reports. Most notably, the Surgeon General's reports in 1980 and 2001 dealt specifically with the health risks of smoking for females.[13] Reports in 1998 and 1994 addressed tobacco use among adolescents.[14]

Important in our most recent era of substance abuse research was a split in the structure and funding of Federal research support, distinguishing between research and treatment (Sloboda 2012). This change, anchored in a new definition of addiction that emerged in the 1980s focusing on behavior (as opposed to unalterable personality characteristics), facilitated a shift of professional focus on an emerging addiction treatment enterprise, fostering development of new treatment approaches and applied research examining the success and dissemination of these new initiatives. Until the early 1990s, and consistent with the original "three-legged stool" approach supporting research, training, and services, addiction services and research were principally funded through the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA), which was established in 1973 as the successor to the NIMH (Sloboda 2012). However, in 1992, bowing to inevitable tensions associated with this arrangement, the research components were moved to NIH and the service components were organized under the newly establish Substance Abuse and Mental Health Services Administration (SAMHSA), whose mission was to lead public health efforts to advance behavioral health, such as reducing the impact of substance abuse and mental illness on

[11]Prior to the establishment of NIAAA, research on alcohol addiction was conducted within the National Institute of Mental Health (NIMH). Since 1974, NIAAA has been an independent Institute of the National Institutes of Health (Warren and Hewitt 2010).

[12]NIDA supports research on nicotine addiction and funds some studies of cessation programs. The CDC's Office of Smoking and Health is the lead Federal agency for comprehensive tobacco prevention and control. The AHRQ supports Cochrane Collaboration Reviews, as well as systematic reviews and meta-analyses.

[13]In 2001, Surgeon General David Satcher stated that, "Women not only share the same health risk as men, but are also faced with health consequences that are unique to women, including pregnancy complications, problems with menstrual function, and cervical cancer."

[14]Overall, expansion of addiction research occurred in recent decades to include a broader range of target at-risk populations than previously studied, including women, veterans, homeless, LGBTI populations, and the elderly.

America's communities. SAMHSA was also charged with conducting applied research to evaluate service programs, as well as the support of ongoing national surveillance systems, such as the National Survey on Drug Use and Health (Sloboda 2012).[15]

Prevention also became a major focus in the 1980s, with the prevention branch at NIDA created in 1982 within the Division of Epidemiology and Prevention Research (Cázares and Beatty 1994; Bell and Battjes 1985). This NIDA Division supports research on the development, testing, and implementation of prevention interventions across a variety of contexts, including early intervention and drug abuse prevention services research, systems and methodology research, along with a full range of drug abuse epidemiology (including nicotine addiction), and services research.[16] Similar structures exist within NIAAA, charged with reducing alcohol-related mortality and morbidity and other alcohol-related problems and consequences through the integration and application of epidemiology and prevention science. SAMHSA also took a leading role in pioneering efforts to catalog knowledge and evaluate the application of evidence-based prevention with aims to further policy and program development (Brounstein et al. 2006; Condon et al. 2008; Marsh et al. 1996). Prevention research, consequently, emerged as a new area of inquiry with its own unique methodological challenges (Botvin 2004; Brounstein et al. 1997; Cázares 1994; Elliott and Mihalic 2004; Hersch et al. 2000; Renes et al. 2007). Also strengthened was the focus on health services research related to both substance abuse prevention and treatment, as efforts were made to improve translation of research into practice, progress in improving community-based care, as well as efforts to offset the social costs of addiction (Compton et al. 2005; Delany et al. 2008). Private foundation funding, such as the Robert Wood Johnson Foundation's support of policy, prevention, and treatment research/programs, also played an important role in advancing substance abuse research opportunities, distinguishing funding along similar lines.

Coinciding with the expansion of resources dedicated to addiction research in recent decades, the nature of investigation also fundamentally changed, largely by way of methodological advances in laboratory, clinical, and community-based research. The reviews conducted by the Surgeon General's Office, for instance, were in and of themselves, methodological advances, serving as precursors to today's complex comprehensive systematic reviews and meta-analyses (Samet 2016). Transformations were also noted in laboratory and clinical research, almost too numerous to mention. An early change was the discontinuation of prisoner research in the early 1970s.[17] This development spurred the creation of new loci of addiction research beyond Lexington and the ARC, the latter of which moved to Baltimore in 1979 to become part of NIDA (Campbell 2010; Kreek et al. 2004). These new centers and institutes collaborated with Federal agencies, such as NIAAA and NIDA, in development and application of innovative research strategies to better understand addiction, including research on special populations (Compton et al. 2005;

[15]State governments also support research; however, most of this funding also comes from the Federal government, which is passed through state agencies.

[16]Within the NIDA Division of Epidemiology and Prevention Research, areas of emphasis include, but are not limited to (1) development of new theoretical approaches to epidemiology, services, and prevention research, (2) determination of intrapersonal, environmental, and genetic factors important in the development of drug abuse/addictions, and (3) development of effective strategies to ensure that evidence-based practices are optimally utilized in the development of services to prevent and treat drug abuse/addictions (Cázares 1994).

[17]While conclusion of prisoner studies is often linked to the release of the Tuskegee Report in 1972, research in prisons was still possible. It did become increasingly difficult, however, following the American Correctional Association's (ACA) release of its first informed consent protocol for correctional institutions in 1972 and the placing of prisoners in the category of vulnerable dependents (Campbell 2010). The ACA later moved to eliminate prison research entirely by withholding accreditation from facilities where it was conducted (Campbell 2010).

Leshner 2000; Millstein 1994; NIDA 2015; Thomas and Conway 2010).

Other changes were directly linked to improved technology. Clinical studies, for instance, were enriched by the proliferation of increasingly sophisticated electronic medical record systems, which allowed researchers to better track subjects and understand drug response (Roden et al. 2012). Increasingly sensitive and specific measurement and analytical techniques allowed for new pharmacokinetic and metabolism studies in humans, furthering understanding of the absorption, distribution, metabolism, and elimination of drugs (including nicotine) and/or alcohol from the body, as well as their impact. These advances ultimately paved the way for an improved understanding of the molecular and cellular mechanisms/genetics of addiction, with wide implications for both policy and practice, such as treatment strategies tailored for high-risk populations (Benowitz 2008; Berrendero et al. 2010; Gorini et al. 2013; Koob et al. 2004; Koob 2006; Kreek et al. 2004; Litten et al. 2010; Riggs et al. 2007). Traditional research tools were also enhanced; particularly substance use surveys, to better control for the biases associated with the reports of sensitive topics such as substance abuse (Gfroerer et al. 1997; Johnson and Fendrich 2005; Kypri et al. 2004; Meiklejohn et al. 2012; Richardson et al. 2003; Weisner et al. 1995; see also Chap. 13 in this volume). Methodological options were also developed to improve self-report. These options now include innovations such as Audio Computer-Assisted Self Interviews (ACASI), or the use of multimedia, specifically pre-recorded audio, in addition to the on-screen text, to facilitate improved substance use data reporting, including reporting by high-risk populations (Currivan et al. 2004; Gribble et al. 2000; Lessler and O'Reilly 1997; Mullany et al. 2013; Turner et al. 1998).[18]

ACASI has also shown promise in clinical study applications (McNeely et al. 2016; Perlis et al. 2004). Biological measures, including urine, hair, and oral fluid testing have further improved epidemiological studies of addiction, including population-based studies (Cook et al. 1997; Fendrich et al. 2004; Palamar et al. 2016; Turner and Miller 1997; see also Chap. 11 in this volume).[19] Public health surveillance was also enhanced by way of improved sampling and the use of new tools, such as geospatial mapping and cellphone/online data-gathering methods to collect information pertaining to substance use and related community-level factors (Kuntsche and Lebhart 2014; Mazumdar et al. 2015; see also Chap. 12).

Lastly, methodologies for treatment outcome studies were improved, including behavioral research and economic evaluation methods (Robinson et al. 2014; see, for example, Chaps. 12 and 14 in this volume). While not all-encompassing, the list of innovative research examples could go on and on, with advances in one area applicable to research in many other areas of inquiry within the substance use field, including program/intervention assessment and epidemiological studies (Greenfield and Kerr 2008; Leshner 2000).

1.4 Increasing Interdisciplinary and Transdisciplinary Research

One major innovation was the evolution of interdisciplinary and transdisciplinary substance abuse research. Due to its complexity, substance abuse research has spanned many disciplines, including but not limited to pharmacology, medicine, the neurosciences, public policy, and the social and behavioral sciences (Sussman and Unger 2004; Sussman et al. 2004). However, historically, work (as well as methodological traditions) from diverse disciplines has not

[18]Not all studies support the use of ACASI to improve self-report. For example, a study by Fendrich et al. (2005) found self-report sensitivity estimates for tobacco use in a drug use survey to be well below the 90% level. Other studies have noted mixed effects of ACASI (Couper et al. 2003; Gribble et al. 2000; Turner et al. 2005).

[19]According to Fendrich et al. (2004), the utility of testing for surveys depends on both the type of substance being examined and the type of test employed, with multiple tests generally having more utility than a single test.

always been well integrated in substance abuse research (Abrams and Clayton 2001; Westermeyer 1990). While challenging, improved communication, statistical approaches, and technology have facilitated the integration of research in new and innovative ways, resulting in a concerted move toward research crossing disciplinary lines. Very specifically, this trend has supported transdisciplinary research, as scientists sought to move beyond simply recognizing inputs by different disciplines to actively establishing—and building upon—connections across disparate research traditions (Kessel and Rosenfield 2008; Klein 2008). Transdisciplinary research is problem focused, collaborative, and differs significantly from interdisciplinary scholarship in that it is characterized by a full integration of epistemologies in the development of study methodology, effectively breaking down disciplinary boundaries (Wickson et al. 2006).

As compared to more traditional lines of scholarship, teams of transdisciplinary collaborators can advance science as they bring to bear and integrate different theories, methodologies, statistical approaches, data, and research traditions; resulting in better quality science, increased innovation, and accelerated translation of evidence into practice (Bozeman and Corely 2004; Chou et al. 2004).[20] Driving factors in this transition include the recognized need to understand the complex array of individual and contextual factors influencing both the use and misuse of drugs and alcohol (Mermelstein et al. 2007; Turner et al. 2004). This realization extends even to fields such as genetic research, where context remains critical to understanding the mechanisms of addiction, thereby necessitating a broader perspective (Giovino 2002; Turner et al. 2004). Transdisciplinary research perspectives have also been integrated into all areas of inquiry, such as prevention as well as intervention and treatment design and evaluation (see also Chap. 2), particularly due to their ability to support tailored interventions (Alemagno 2009; Baker et al. 2003; Compton et al. 2005; Lieber 1989; Sloboda et al. 1998; Sussman et al. 2004). They have also become an important element in the research agendas of government agencies such as the National Cancer Institute, as well as private foundations such as the Robert Wood Johnson Foundation (Kessel and Rosenfield 2008). These agendas have supported transdisciplinary research/prevention/treatment initiatives as well as new research centers, such as Transdisciplinary Tobacco Use Research Centers (Kobus and Mermelstein 2009; Mermelstein et al. 2007; Turkkan et al. 2000) or the NIDA-funded Transdisciplinary Prevention Research Centers, which supports both research that translates theories to practice and policies that prevent substance use.

Transdisciplinary research has also been a major thematic element at professional conferences, such as the "Reflections on 40 Years of Drug Abuse Research" meeting in Key Largo, Florida in 2006, resulting in a special issue of the *Journal of Drug Issues* (Sloboda et al. 2009a, b). Despite this progress, the promise of a fully transdisciplinary approach to addiction research has not yet been fully realized, and the need for better integration of data systems, theoretical and analytical models, and intentional connections crossing disciplinary silos persists. The latter, in particular, is not easy, as these collaborations require considerable effort and time (Mermelstein et al. 2007; Provan et al. 2008).

1.5 Continued Challenges and New Opportunities

Collectively, the methodological advances in substance abuse research did not happen overnight, and, even today, remain a work in progress. An early comprehensive review of common methodological problems associated with addiction research by Nathan and Lansky (1978) identified a number of ongoing concerns, including selective or biased reviews of the literature, reliance on incomplete diagnostic criteria for study inclusion, inadequately accounting for

[20]The Institute of Medicine has broadly called for a shift to research that engages investigators from multiple fields and disciplines to better capitalize on rapidly expanding knowledge of how genetic, social, and environmental factors impact health (Hernandez and Blazer 2006).

study dropouts, and failure to follow subjects for adequate lengths of time. Even today, despite noted advancements, there remain a number of methodological issues that have yet to be resolved. Surveys, for example, while being the primary source for much of what we know about drug and alcohol abuse, are plagued by methodological failures, including sampling, coverage, nonresponse, measurement, and processing errors (Fendrich et al. 2005; Gfroerer and Kennet 2015; Gfroerer et al. 1997; Giovino 2002; Grucza et al. 2007; Johnson and Fendrich 2005; Johnson 2012, 2014, 2015; Kremling 2013; Midanik et al. 2013; Sevigny and Fuleihan 2015) see also Chap. 13 this volume. As these methodological and conceptual failures continue to hinder understandings of substance abuse, improving the collection and use of data is critical to the value of the information and the conclusions produced (Johnson 2012). Treatment and prevention researchers face similar issues related to study design, principally related to studies utilizing addicts as subjects (Booth and Watters 1994; Flay and Petraitis 1991; Sloboda et al. 1998).

While improving over time, key issues still include intervention exposure/compliance, implementation fidelity, assessment of exposure and outcome measures, sampling attrition, accuracy of subject reports, and the choice of analytic model; necessitating consideration of new and innovative designs, including those incorporating "real-world" contexts of service delivery (Alemagno 2009; Baker et al. 2011; Borders and Booth 2007; Clark and Winters 2002; Colby et al. 2004; Compton et al. 2005; Galea et al. 2004; Robinson et al. 2014; Sloboda et al. 1998, 2009a, b; Willenbring 2010). This includes economic evaluation of substance abuse services and interventions (French and Drummond 2005).

Also problematic are the several views that still exist regarding the etiology of substance use and abuse, each weighing somewhat differently the relative contributions of genetic, individual, cultural, and social influences. Resolution is not yet fully possible, as even recent advancements in neurobiological research on addiction, such as increasingly sensitive and specific analytical techniques, as well as improved information on the contributions of gene variations to vulnerability to addiction, cannot fully articulate all of the factors contributing to addiction across diverse populations (Foroud et al. 2010; Hall et al. 2008; Kalant 2009; Kreek et al. 2004; Obot et al. 2004; Trujillo et al. 2006; Volkow and Baler 2014). Additional research is also needed on the impact of misuse on individual function (Scott et al. 2007). Moreover, despite the promise of neuroscience research, caution is necessary when relying solely on a single explanation, so as to avoid overly deterministic causal models of addiction that mask the complex interaction between environment and individual, again making an argument for more of a transdisciplinary focus with all the inherent challenges therein.

With regard to the transfer of evidence into practice, there have again been noted improvements due, in part, to advances in the research process, which has compelled revisions of best practices implementation, especially with regard to preventive interventions attempting to maximize population impact (Millstein 1994; Sloboda 2014; Spoth et al. 2013).

Lastly, in examining progress made in addiction research, it is important to remember that the issue itself is a moving target, with new and emerging drugs and risk populations. Substance abuse research is also influenced by advances in research methods, further complicating the picture. For example, new research methodologies, such as web mapping, have been used to more rapidly identify new and emerging trends in substance abuse. One example is the Psychonaut Web-Mapping Project, a European collaboration which monitors discussion forums, social media, and other internet resources to rapidly identify emerging trends in novel psychoactive substances warranting public health response (Deluca et al. 2012). Similar web-mapping initiatives, as well as the use of other internet-based open-source tools, have been used elsewhere to better understand and respond to the changing array of emerging psychoactive substances entering the marketplace, as well as other trends in substance abuse (Brownstein et al.

2009; Bruno et al. 2013; Butler et al. 2007; Young et al. 2015).[21]

Each new trend that is discovered, in turn, creates its own methodological challenges for scientists seeking understanding and/or solutions, as shifts often involve unique risk populations as defined by geography, age, culture, socioeconomic status, and the like. Subsequent solutions take many forms, including the leveraging of new technologies and alternate forms of communication, which is evident in smartphone and other handheld technologies that have opened up opportunities for the assessment of substance use/misuse via text messaging and other forms of electronic contact, as well as providing for new intervention opportunities (Bernhardt et al. 2007; Kuntsche and Labhart 2012, 2014; Kuntsche and Robert 2009; Phillips et al. 2014; Sufoletto et al. 2012). This includes important advances in the collection of ecological momentary assessment (EMA) data on use, as well as daily factors associated with the abuse of substances such as alcohol and tobacco (Collins et al. 2003; Freedman et al. 2006; Minami et al. 2010; Shiffman 2009). While still an emerging area that is not without its challenges in substance abuse research, available studies have found these assessments to be both feasible and valid (Collins et al. 2003; Ferguson and Shiffman 2011; Galloway et al. 2008; Phillips et al. 2014; Serre et al. 2012; Shiffman 2009). Today, these technological advances, including the aforementioned GPS technologies, are increasingly used to collect real-time use and behavioral data associated with the use and abuse of alcohol, drugs, and tobacco. Overall improvements in informatics generally, as well as team science, have facilitated further integration of these data; allowing for more rapid analysis across diverse substances and risk populations.

Ethical concerns have been expressed related to the use of real-time data, but it remains possible to remove identifiers from the data and use of passwords can further reduce the risk of privacy violations (Beckjord and Shiffman 2014). The use of handheld technology is illustrative of new tool utilization for improved study design, including studies identifying or responding to an emergent trend. This cycle is ongoing, regardless of the substances under investigation, as the field constantly seeks to improve data that can be effectively utilized to more rapidly inform interventions as new problems arise.

[21]Web-based bioinformatics and other open-source research and development are also being used to spur drug discovery and assessment (Chen and Butte 2016; Wishart 2005).

1.6 Conclusion

The history of substance abuse research and treatment is complex, spanning multiple disciplines, each with their own research traditions. It also involves multiple substances, both alone and in combination. Still, advances in the methodologies employed in terms of conceptual sophistication, study design, measurement, and data analysis have built upon one another in a transdisciplinary manner; greatly expanding our knowledge of the mechanisms of addiction, as well as informing new and innovative solutions, including efforts to prevent abuse altogether (O'Brien 2003; Sloboda 2014; Treno et al. 2014). This has been especially true over the last several decades, as neurogenetic research methods and models, coupled with advances in technology and bioinformatics, have the potential to finally resolve, or at least reconcile, competing explanatory models of addiction that have dominated the scientific debate; historically defining addiction as either principally stemming from an individual's moral or medical state. Caution is necessary, however, lest we fall into the historical trap of trying to establish more simplified linear causal models, as there are inherent limitations to any scientific discipline. Despite the noted advances in neuroscience, addiction remains a behavioral disorder generated within exceedingly complex interactions of agent (e.g., drug, alcohol, or tobacco), user, and environment (Kalant 2009). The social sciences remain essential to our understanding of abuse, helping explain the need, mechanisms of distribution (access), economic implications

of addiction, and the like. There are certain aspects of addiction that simply cannot be explained by neurogenetic research. As elegantly stated by Kalant, "This [*sic*] is no longer the terrain of pharmacology or neurobiology or psychology or sociology, but an amalgam of them all" (2009). A transdisciplinary perspective is foundational to further progress.

There are other important considerations when assessing past progress in addiction research as a marker for future development. It is important, for instance, to remember that progress has been nonlinear; shaped, in part, by larger political and social forces. For instance, financial support for drug and alcohol research over the years has been impacted by a number of factors (Musto 1996; Sloboda 2012). Generally, when drug or alcohol abuse has been viewed as a major crisis, money flows in support of addiction research, thus spurring advances. However, support often wanes when abuse rates stabilize or shift away from high-priority substances. Additionally, as progress is made, especially with regard to the neuroscientific and genetic elucidation of the mechanisms of addiction, scientists must increasingly anticipate the ethical issues that arise from this work to identify individual biomarkers for risk, including the capacity of addicted persons to give consent to treatment, individual privacy, and risk of coercion (Hall et al. 2004).

Lastly, it is important to remember that much remains to be known. Most methodological advances in addiction research have largely occurred only in the last 50 years. This is especially true of research on key risk groups, such as women, children, LGBTI populations, and the elderly, that have often been historically neglected, both in the U.S. and globally. Also challenging is the noted fluidity of the subject matter, with new and emerging substances and risk populations changing constantly. As a result, there are calls for an urgent need to review and improve the quality and timeliness of substance abuse data, its implications, and intervention outcomes, theoretically facilitating an effective clinical and public health response (Degenhardt et al. 2011; Fischbein and Ridenour 2013; French and Drummond 2005; Gowing et al. 2015; Riggs et al. 2007).

Acknowledgements The authors thank Drs. Peggy and Richard Stephens for their helpful comments on earlier drafts of the chapter.

References

Abbott, R., & Smith, D. E. (2015). The new designer drug wave: A clinical, toxicological, and legal analysis. *Journal of Psychoactive Drugs, 47,* 368–371.

Abrams, D. B., & Clayton, R. R. (2001). Transdisciplinary research to improve brief interventions for addictive behaviors. In P. M. Monti, S. M. Colby, & T. A. O'Leary (Eds.), *Adolescents, alcohol, and substance abuse: Reaching teens through brief interventions* (pp. 321–341). New York, NY: Guilford Press.

Acker, C. J. (1995). Addiction and the laboratory: The work of the National Research Council's Committee on Drug Addiction, 1928–1939. *Isis*, *86*, 167–193.

Acker, C. (2002). *Creating the American Junkie: Addiction research in the classic era of narcotics control*. Baltimore, MD: Johns Hopkins University Press.

Acker, C. J. (1997). The early years of the PHS narcotic hospital at Lexington, Kentucky. *Public Health Reports, 112,* 245–247.

Advisory Committee to the Surgeon General of the Public Health Service. (1964). *Smoking and health.* PHS Publication No. 1103. Washington, DC: Department of Health, Education, and Welfare.

Agaku, I. T., King, B. A., Husten, C. G., Bunnell, R., Ambrose, B. K., Hu, S., et al. (2014). Tobacco product use among adults—United States, 2012–2013. *Morbidity and Mortality Weekly Report, 63*, 542–547.

Alemagno, S. A. (2009). Drug abuse research: A shifting paradigm. *Journal of Drug Issues, 39*(1), 223–226.

American Society of Addiction Medicine. (2015). *Increased marijuana, heroin use contribute to highest reported illicit drug use in more than a decade*. Chevy Chase, MD: ASAM.

Baker, T. B., Hatsukami, D. K., Lerman, C., O'Malley, S. S., Shields, A. E., & Fiore, M. C. (2003). Transdisciplinary science applied to the evaluation of treatments for tobacco use. *Nicotine & Tobacco Research, 5,* S89–S99.

Baker, T. B., Mermelstein, R., Collins, L. M., Piper, M. E., Jorenby, D. E., Smith, S. S., et al. (2011). New methods for tobacco dependence treatment research. *Annals of Behavioral Medicine, 41,* 192–207.

Beckjord, E., & Shiffman, S. (2014). Background for real-time monitoring and interventions related to alcohol use. *Alcohol Research, 36,* 9–18.

Bell, C. S., & Battjes, R. J. (1985). Overview of drug abuse prevention research. *NIDA Research Monograph, 63,* 1–7.

Benowitz, N. L. (2008). Neurobiology of nicotine addiction: Implications for smoking cessation and treatment. *American Journal of Medicine, 121*(4A), S3–S10.

Bernhardt, J. M., Usdan, S., Mays, D., Arriola, K. J., Martin, R. J., Cremeens, J., et al. (2007). Alcohol assessment using wireless handheld computers: A pilot study. *Addictive Behavior, 32,* 3065–3070.

Berrendero, F., Robledo, P., Trigo, J. M., Martin-Garcia, E., & Maldonado, R. (2010). Neurobiological mechanisms involved in nicotine dependence and reward: Participation of the endogenous opioid system. *Neuroscience and Biobehavioral Reviews, 35,* 220–231.

Booth, R. E., & Watters, J. K. (1994). How effective are risk-reduction interventions targeting injection drug users? *AIDS, 8,* 1515–1524.

Borders, T. F., & Booth, B. M. (2007). Research on rural residence and access to drug abuse services: Where are we and where do we go? *Journal of Rural Health, 23*(Suppl 1), 79–83.

Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. *Prevention Science, 5*(1), 69–72.

Bouchery, E. E., Harwood, H. J., Sacks, J. J., Simon, C. J., & Brewer, R. D. (2011). Economic costs of excessive alcohol consumption in the US, 2006. *American Journal of Preventive Medicine, 41*(5), 516–524.

Bozeman, B., & Corely, E. (2004). Scientists' collaboration strategies: Implications for scientific and human capital. *Research Policy, 33,* 599–616.

Brounstein, P. J., Emshoff, J. G., Hill, G. A., & Stoil, M. J. (1997). Assessment of methodological practices in the evaluation of alcohol and other drug (AOD) abuse prevention. *Journal of Health & Social Policy, 9*(2), 1–19.

Brounstein, P. J., Gardner, S. E., & Backer, T. E. (2006). Research to practice: Efforts to bring effective prevention to every community. *Journal Primary Prevention, 27*(1), 91–109.

Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—Harnessing the web for public health surveillance. *New England Journal of Medicine, 360,* 2153–2157.

Bruno, R., Poesiat, R., & Matthews, A. J. (2013). Monitoring the internet for emerging psychoactive substances available in Australia. *Drug and Alcohol Review, 32,* 541–544.

Burgheim, R. A. (1953). Yale Center of Alcohol Studies investigates drinking habits of carefree undergraduates: Find college life affects drinking little; Habits acquired earlier. Harvard Crimson, November 21. Accessed: http://www.thecrimson.com/article/1953/11/21/yale-center-of-alcohol-studies-investigates/

Butler, S. F., Venuti, S. W., Benoit, C., Beaulaurier, R. L., Houle, B., & Katz, N. (2007). Internet surveillance: Content analysis and monitoring of product-specific internet prescription opioid abuse-related postings. *Clinical Journal of Pain, 23,* 619–628.

Campbell, N. D. (2006). "A new deal for the drug addict": The Addiction Research Center, Lexington, Kentucky. *Journal of the History of the Behavioral Sciences, 42,* 135–157.

Campbell, N. D. (2007). *Discovering addiction: The science and politics of substance abuse research*. Ann Arbor, MI: University of Michigan Press.

Campbell, N. D. (2010). The history of a public science: How the Addiction Research Center became the NIDA intramural research program. *Drug and Alcohol Dependence, 107,* 108–112.

Candon, P. M., Ward, J. H., & Pandina, R. J. (2014). The journal of studies on alcohol and drugs and the Rutgers Center of Alcohol Studies: A history of the evolution of alcohol research. *Journal of Studies on Alcohol and Drugs, 75*(Suppl 17), 8–17.

Cázares, A. (1994). Prevention intervention research: Focus and perspective. In A. Cázares & L. A. Beatty (Eds.), *Scientific methods for prevention research*. NIDA Research Monograph 139.

Cázares, A., & Beatty, L. A. (1994). Introduction: Scientific methods for prevention intervention research. In A. Cázares & L.A. Beatty (Eds.) *Scientific methods for prevention research*. NIDA Research Monograph 139.

Center for Behavioral Health Statistics and Quality. (2015). Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health. HHS Publication No. SMA 15-4927, NSDUH Series H-50. Retrieved from http://www.samhsa.gov/data/

Centers for Disease Control and Prevention. (1999). Achievements in public health, 1900–1999: Tobacco use – United States, 1900–1999. *Morbidity and Mortality Weekly Report*, *48*, 986–993.

Chen, B., & Butte, A. J. (2016). Leveraging big data to transform target selection and drug discovery. *Clinical Pharmacology and Therapeutics, 99,* 285–297.

Chou, C. P., Spruijt-Metz, D., & Azen, S. P. (2004). How can statistical approaches enhance transdisciplinary study of drug misuse prevention? *Substance Use and Misuse, 39*(10–12), 1867–1906.

Clark, D. B., & Winters, K. C. (2002). Measuring risks and outcomes in substance use disorders prevention research. *Journal of Consulting and Clinical Psychology, 70,* 1207–1223.

Colby, S. M., Lee, C. S., Lewis-Esquerre, J., Esposito-Smythers, C., & Monti, P. M. (2004). Adolescent alcohol misuse: Methodological issues for enhancing treatment research. *Addiction, 99,* 47–62.

Collins, R. L., Kashdan, T. B., & Gollnisch, G. (2003). The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. *Experimental and Clinical Psychopharmacology, 11,* 73–78.

Compton, W. M., Stein, J. B., Robertson, E. B., Pintello, D., Pringle, B., & Volkow, N. D. (2005). Charting the course of health services research at the National

Institute on Drug Abuse. *Journal of Substance Abuse Treatment, 29*(3), 167–172.

Compton, W. M., & Volkow, N. D. (2006). Major increases in opioid analgesic abuse in the United States: Concerns and strategies. *Drug and Alcohol Dependence, 81*(2), 103–107.

Condon, T. P., Miner, L. L., Balmer, C. W., & Pintello, D. (2008). Blending addiction research and practice: Strategies for technology transfer. *Journal of Substance Abuse Treatment, 35*(2), 156–160.

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly, 67*(3), 385–395.

Cummings, M. K. (2002). Programs and policies to discourage the use of tobacco products. *Oncogene, 21,* 7349–7364.

Cummings, M. K., Brown, A., & O'Conner, R. (2007). The cigarette controversy. *Cancer Epidemiology, Biomarkers and Prevention, 16*(6), 1070–1076.

Currivan, D., Nyman, A. L., Turner, C. F., & Biener, L. (2004). Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly, 68*(4), 542–564.

Dart, R. C., Surratt, H. L., Cicero, T. J., Parrino, M. W., Severtson, S. G., Bucher-Bartelson, B., et al. (2015). Trends in opioid analgesic abuse and mortality in the United States. *New England Journal of Medicine, 372,* 241–248.

Degenhardt, L., Bucello, C., Calabria, B., Nelson, P., Roberts, A., Hall, W., et al. (2011). What data are available on the extent of illicit drug use and dependence globally? Results of four systematic reviews. *Drug and Alcohol Dependence, 117*, 85–101.

Degenhardt, L., & Hall, W. D. (2015). The impact of illicit drugs on public health. In H. H. Brownstein (Ed.), *The handbook of drugs and society*. West Sussex, UK: Wiley.

Degenhardt, L., Whiteford, H., & Hall, W. D. (2014). The global burden of disease projects: What we learned about illicit drug use and dependence and their contribution to the global burden of disease? *Drug and Alcohol Review, 33,* 4–12.

Delany, P. J., Shields, J. J., Willenbring, M. L., & Huebner, R. B. (2008). Expanding the role of health services research as a tool to reduce the public health burden of alcohol use disorders. *Substance Use & Misuse, 43*(12–13), 1729–1746.

Deluca, P., Davey, Z., Corazza, O., Di Furia, L., Farre, M., Holmefjord Flesland, L., et al. (2012). Identifying emerging trends in recreational drug use: Outcomes from the Psychonaut Web Mapping Project. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 39,* 221–226.

Doll, R. (1998). Uncovering the effects of smoking: Historical perspective. *Statistical Methods in Medical Research, 7,* 87–117.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*(1), 47–53.

Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A., & Spiehler, V. (2004). The utility of drug testing in epidemiological research: Results from a general population survey. *Addiction, 99,* 197–208.

Fendrich, M., Mackesy-Amiti, M. E., Johnson, T. P., Hubbell, A., & Wislar, J. S. (2005). Tobacco-reporting validity in an epidemiological drug-use survey. *Addictive Behaviors, 30*(1), 175–181.

Ferguson, S. G., & Shiffman, S. (2011). Using the methods of ecological momentary assessment in substance dependence research—Smoking cessation as a case study. *Substance Use & Misuse, 46,* 87–95.

Fillmore, K. M., & Marden, P. M. (1977). Longitudinal research at the Rutgers Center for Alcohol Studies. *Alcoholism, Clinical and Experimental Research, 1* (3), 251–257.

Fischbein, D. H., & Ridenour, T. A. (2013). Advancing transdisciplinary translation for prevention of high-risk behaviors: Introduction to the special issue. *Prevention Science*, *14*, 201–205.

Flay, B. R., & Petraitis, J. (1991). Methodological issues in drug use prevention research: Theoretical foundations. *NIDA Research Monograph, 107,* 81–109.

Foroud, T., Edenberg, H. J., & Crabbe, J. C. (2010). Genetic research: Who is at risk for alcoholism? *Alcohol Research and Health, 33,* 64–75.

Freedman, M. J., Lester, K. M., McNamara, C., Milby, J. B., & Schumacher, J. E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *Journal of Substance Abuse Treatment, 30,* 105–111.

French, M. T., & Drummond, M. (2005). A research agenda for economic evaluation of substance abuse services. *Journal of Substance Abuse Treatment, 29,* 125–137.

Galea, S., Nandi, A., & Vlahov, D. (2004). The social epidemiology of substance use. *Epidemiologic Reviews, 26,* 36–52.

Galloway, G. P., Didier, R., Garrison, K., & Mendelson, J. (2008). Feasibility of ecological momentary assessment using cellular telephones in methamphetamine dependent subjects. *Substance Abuse, 1,* 9–14.

Gfroerer, J., & Kennet, J. (2015). Collecting survey data on sensitive topics: Substance use. In T. P. Johnson (Ed.), *Handbook of health survey methods*. Hoboken, NJ: Wiley.

Gfroerer, J., Lessler, J., & Parsley, T. (1997). Studies of nonresponse and measurement error in the national household survey on drug abuse. *NIDA Research Monograph, 167,* 273–295.

Giovino, G. A. (2002). The epidemiology of tobacco use in the United States. *Oncogene, 21,* 7326–7340.

Gorini, G., Roberts, A. J., & Mayfield, R. D. (2013). Neurobiological signatures of alcohol dependence revealed by protein profiling. *PLoS ONE, 8*(12), e82656.

Gowing, L. R., Ali, R. L., Allsop, S., Marsden, J., Turf, E. E., West, R., et al. (2015). Global statistics on addictive behaviors: 2014 status report. *Addiction, 110,* 904–919.

Greenfield, T. K., & Kerr, W. C. (2008). Alcohol measurement methodology in epidemiology: Recent advances and opportunities. *Addiction, 103,* 1082–1099.

Gribble, J. N., Miller, H. G., Cooley, P. C., Catania, J. A., Pollack, L., & Turner, C. F. (2000). The impact of T-ACASI interviewing on reported drug use among men who have sex with men. *Substance Use & Misuse, 35*(6–8), 869–890.

Griffiths, P., & Meacham, M. (2008). Illicit drug trends globally. In K. Heggenhougen & S. Quah (Eds.), *International encyclopedia of public health* (pp. 515–523). San Diego, CA: Academic Press.

Grucza, R. A., Abbacchi, A. M., Przybeck, T. R., & Gfroerer, J. C. (2007). Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. *Addiction, 102,* 623–629.

Hall, W., Carter, L., & Morley, K. I. (2004). Neuroscience research on the addictions: A prospectus for future ethical and policy analysis. *Addictive Behaviors, 29,* 1481–1495.

Hall, W. D., Gartner, C. E., & Carter, A. (2008). The genetics of nicotine addiction liability: Ethical and social policy implications. *Addiction, 103,* 350–359.

Hansen, R. N., Oster, G., Edelsberg, J., Woody, G., & Sullivan, S. D. (2011). Economic costs of nonmedical use of prescription opioids. *Clinical Journal of Pain, 3,* 194–202.

Hernandez, L. M., & Blazer, D. G. (2006). *Genes, behavior and the social environment: Moving beyond the nature/nurture debate*. Washington, DC: The National Academies Press.

Hersch, R. K., Cook, R. F., Deitz, D. K., & Trudeau, J. V. (2000). Methodological issues in workplace substance abuse prevention research. *Journal of Behavioral Health Services and Research, 27*(2), 144–151.

Israel, Y., & Lieber, C. S. (2002). The Research Society on Alcoholism. *Addiction, 97,* 483–486.

Jellinek, E. M. (1960). *The disease concept of alcoholism*. New Haven, CT: Hillhouse.

Johnson, T., & Fendrich, M. (2005). Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of Epidemiology, 15*(5), 381–389.

Johnson, T. P. (2012). Failures in substance use surveys. *Substance Use & Misuse, 47*(13–14), 1675–1682.

Johnson, T. P. (2014). Sources of error in substance use prevalence surveys. *International Scholarly Research Notices, 2014,* 21.

Johnson, T. P. (2015). Measuring substance use and misuse via survey research: Unfinished business. *Substance Use & Misuse, 50,* 1134–1138.

Kalant, H. (2009). What neurobiology cannot tell us about addiction. *Addiction, 105,* 780–789.

Kessel, F., & Rosenfield, P. L. (2008). Toward transdisciplinary research: Historical and contemporary perspectives. *American Journal of Preventive Medicine, 35,* S225–S234.

King, B. A., Dube, S. R., & Tynan, M. A. (2012). Current tobacco use among adults in the United States: Findings from the National Adult Tobacco Survey. *American Journal of Public Health, 102,* e93.

Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine, 35,* S116–S123.

Kobus, K., & Mermelstein, R. (2009). Bridging basic and clinical science with policy studies: The partners with transdisciplinary tobacco use research centers experience. *Nicotine & Tobacco Research, 11,* 467–474.

Koob, G. F. (2006). The neurobiology of addiction: A neuroadaptational view relevant for diagnosis. *Addiction, 101*(Suppl 1), 23–30.

Koob, G. F., Ahmed, S. H., Boutrel, B., Chen, S. A., Kenny, P. J., Markou, A., et al. (2004). Neurobiological mechanisms in the transition from drug use to drug dependence. *Neuroscience & Biobehavioral Reviews, 27,* 739–749.

Kreek, M. J., Schlussman, S. D., Bart, G., LaForge, K. S., & Butelman, E. R. (2004). Evolving perspectives on neurobiological research on the addictions: Celebration of the 30th anniversary of NIDA. *Neuropharmacology, 47*(Suppl 1), 324–344.

Kremling, J. (2013). *Estimating drug use: The effect of sampling methods and policy implications*. El Paso, TX: LFB Scholarly Publishing LLC.

Kuntsche, E., & Lebhart, F. (2012). Investigating the drinking patterns of young people over the course of the evening at weekends. *Drug and Alcohol Dependence, 124,* 319–324.

Kuntsche, E., & Lebhart, F. (2014). The future is now—Using personal cellphones to gather data on substance use and related factors. *Addiction, 109,* 1052–1053.

Kuntsche, E., & Robert, B. (2009). Short message service (SMS) technology in alcohol research—A feasibility study. *Alcohol and Alcoholism, 44,* 423–428.

Kypri, K., Stephenson, S., & Langley, J. (2004). Assessment of nonresponse bias in an internet survey of alcohol use. *Alcoholism-Clinical & Experimental Research, 28*(4), 630–634.

Lankenau, S., Schrager, S. M., Silva, K., Kecojevic, A., Bloom, J. J., Wong, C., et al. (2012). Misuse of prescription and illicit drugs among high-risk young adults in Los Angeles and New York. *Journal of Public Health Research, 1*, e6.

Lee, Y. O., Hebert, C. J., Nonnemaker, J. M., & Kim, A. E. (2014). Multiple tobacco product use among adults in the United States: Cigarettes, cigars, electronic cigarettes, hookah, smokeless tobacco, and snus. *Preventive Medicine, 62,* 14–19.

Leshner, A. I. (2000). National Institute on Drug Abuse's behavioral research agenda. *Experimental and Clinical Psychopharmacology, 8*(3), 273–275.

Lessler, J. T., & O'Reilly, J. M. (1997). Mode of interview and reporting of sensitive issues: Design and implementation of audio computer-assisted

self-interviewing. *NIDA Research Monograph, 167,* 366–382.

Lieber, C. S. (1989). NIAAA and alcohol research. *Alcohol Health and Research World, 13*(4), 301–304.

Lipton, D. S., & Johnson, B. D. (1998). Smack, crack and score: Two decades of NIDA-funded drugs and crime research at NDRI 1974–1994. *Substance Use & Misuse, 33*(9), 1779–1815.

Litten, R. Z., Bradley, A. M., & Moss, H. B. (2010). Alcohol biomarkers in applied settings: Recent advances and future research opportunities. *Alcoholism, Clinical and Experimental Research, 34*(6), 955–967.

Lowney, J. (2008). *Stoned, drunk, or sober? Understanding alcohol and drug use through qualitative, quantitative, and longitudinal research*. Lanham, MD: University Press of America Inc.

MacKinnon, D. P., & Lockwood, C. M. (2003). Advances in statistical methods for substance abuse prevention research. *Prevention Science, 4,* 155–171.

Marsh, A., Jansen, M., Lewis, C., & Straw, R. B. (1996). Evaluation in the substance abuse and Mental Health Services Administration. *Evaluation and the Health Professions, 19*(3), 363–376.

Maxwell, J. C. (2011). The prescription drug epidemic in the United States: A perfect storm. *Drug and Alcohol Review, 30,* 264–270.

May, E. L., & Jacobson, A. E. (1989). The committee on problems of drug dependence: A legacy of the National Academy of Sciences. A historical account. *Drug and Alcohol Dependence, 23,* 183–218.

Mazumdar, S., Mcrae, I. S., & Islam, M. M. (2015). How can geographical information systems and spatial analysis inform a response to prescription opioid misuse? A discussion in the context of existing literature. *Current Drug Abuse Reviews, 8*(2), 104–110.

McMillen, R., Maduka, J., & Winickoff, J. (2012). Use of emerging tobacco products in the United States. *Journal of Environmental and Public Health, 2012,* 989474. doi:10.1155/2012/989474.

McNeely, J., Strauss, S. M., Rotrosen, J., Ramautar, A., & Gourevitch, M. N. (2016). Validation of an audio computer-assisted self-interview (ACASI) version of the alcohol, smoking and substance involvement screening test (ASSIST) in primary care patients. *Addiction, 111*(2), 233–244.

Meiklejohn, J., Connor, J., & Kypri, K. (2012). The effect of low survey response rates on estimates of alcohol consumption in a general population survey. *PLoS ONE, 7*(4), e35527.

Mermelstein, R., Kobus, K., & Clayton, R. (2007). Transdisciplinary tobacco use research: A decade of progress. *Nicotine & Tobacco Research, 9*(Suppl 4), S519–S522.

Midanik, L. T., Ye, Y., Greenfield, T. K., & Kerr, W. (2013). Missed and inconsistent classification of current drinkers: Results from the 2005 US National Alcohol Survey. *Addiction, 108,* 348–355.

Millstein, R. A. (1994). Drug abuse research accomplishments and opportunities: A report from the national perspective. *NIDA Research Monograph, 140,* 11–22.

Minami, H., McCarthy, D. E., Jorenby, D. E., & Baker, T. B. (2010). An ecological momentary assessment analysis of relations among coping, affect and smoking during a quit attempt. *Addiction, 106,* 641–650.

Mullany, B., Barlow, A., Neault, N., Billy, T., Hastings, R., Coho-Mescal, V., et al. (2013). Consistency in reporting of sensitive behaviors by adolescent American Indian women: A comparison of interviewing methods. *American Indian and Alaska Native Mental Health Research, 20*(2), 42–51.

Musto, D. F. (1996). Drug abuse research in historical perspective. In Committee on Opportunities in Drug Abuse Research (Eds.), *Pathways of addiction: Opportunities in drug abuse research*. Washington, DC: National Academy Press.

Musto, D. F. (1997). Historical perspectives. In J. H. Lowinson et al. (Eds.), *Substance abuse: A comprehensive textbook* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Nathan, P. E., & Lansky, D. (1978). Common methodological problems in research on the addictions. *Journal of Consulting and Clinical Psychology, 46* (4), 713–726.

National Commission on Marihuana and Drug Abuse. (1973). *Drug use in America: Problem in perspective*. Washington, DC: U.S. Government Printing Office.

National Institute on Drug Abuse. (1995). History of the Addiction Research Center. *NIDA Notes, 10*(6). https://archives.drugabuse.gov/NIDA_Notes/NNVol10N6/ARCHistory.html

Neale, J., Allen, D., & Coombes, L. (2005). Qualitative research methods within the addictions. *Addiction, 100,* 1584–1593.

NIAAA. (2011). NIAAA: 40 years of research leadership. *Alcohol Alert*, 79. http://pubs.niaaa.nih.gov/publications/AA79/AA79.htm

NIDA. (2015). National Institute on Drug Abuse 2016–2020 Strategic Plan. https://www.drugabuse.gov/sites/default/files/2016-2020nidastrategicplan.pdf

NIDA. (2016). Trends & Statistics. https://www.drugabuse.gov/related-topics/trends-statistics#costs. Accessed April 2016.

Obot, I. S., Poznyak, V., & Monteiro, M. (2004). From basic research to public health policy: WHO report on the neuroscience of substance dependence. *Addictive Behaviors, 29,* 1497–1502.

O'Brien, C. P. (2003). Research advances in the understanding and treatment of addiction. *American Journal on Addictions, 12,* S36–S47.

Page, P. B. (1988). The origins of alcohol studies: E.M. Jellinek and the documentation of the alcohol research literature. *British Journal of Addiction, 83,* 1095–1103.

Page, P. B. (1997). E.M. Jellinek and the evolution of alcohol studies: A critical essay. *Addiction, 92*(12), 1619–1637.

Palamar, J. D., & Acosta, P. (2015). Synthetic cannabinoid use in a nationally representative sample of US high school seniors. *Drug and Alcohol Dependence, 149,* 194–202.

Palamar, J. J., Martins, S. S., Su, M. K., & Ompad, D. C. (2015). Self-reported use of novel psychoactive substances in a US nationally representative survey: Prevalence, correlates, and a call for new survey methods to prevent underreporting. *Drug and Alcohol Dependence, 156,* 112–119.

Palamar, J. J., Salomone, A., Vincenti, M., & Cleland, C. M. (2016). Detection of "bath salts" and other novel psychoactive substances in hair samples of ecstasy/MDMA/"Molly" users. *Drug and Alcohol Dependence, 161,* 200–205.

Parascandola, M. (2001). Cigarettes and the US Public Health Service in the 1950s. *American Journal of Public Health, 91*(2), 196–205.

Parascandola, M. (2004). Skepticism, statistical methods, and the cigarette: A historical analysis of a methodological debate. *Perspectives in Biology and Medicine, 47*(42), 244–261.

Paulozzi, L. J., Jones, C. M., Mack, K. A., & Rudd, R. A. (2011). Vital signs: Overdoses of prescription opioid pain relievers—United States, 1999–2008. *Morbidity and Mortality Weekly Report, 60*, 1487–1492.

Perlis, T. E., Des Jarlais, D. C., Friedman, S. R., Arasteh, K., & Turner, C. F. (2004). Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. *Addiction, 99*(7), 885–896.

Phillips, M. M., Phillips, K. T., Lalonde, T. L., & Dykema, K. R. (2014). Feasibility of text messaging for ecological momentary assessment of marijuana use in college students. *Psychological Assessment, 26,* 947–957.

Proctor, R. N. (2012). The history of the discovery of the cigarette-lung cancer link: Evidentiary traditions, corporate denial, global toll. *Tobacco Control, 21*(2), 87–91.

Provan, K. G., Clark, P. I., & Huerta, T. (2008). Transdisciplinarity among tobacco harm-reduction researchers: A network analytic approach. *American Journal of Preventive Medicine, 35,* S173–S181.

Renes, S. L., Ringwalt, C., Clark, H. K., & Hanley, S. (2007). Great minds don't always think alike: The challenges of conducting substance abuse prevention research in public schools. *Journal of Drug Education, 37*(2), 97–105.

Richardson, J., Fendrich, M., & Johnson, T. P. (2003). Neighborhood effects on drug reporting. *Addiction, 98* (12), 1705–1711.

Riggs, P. D., Thompson, L. L., Tapert, S. F., Frascella, J., Mikulich-Gilbertson, S., Dalwani, M., et al. (2007). Advances in neurobiological research related to interventions in adolescents with substance use disorders: Research to practice. *Drug and Alcohol Dependence, 91,* 306–311.

Rinaldi, R. C., Steindler, E. M., Wilford, B. B., & Goodwin, D. (1988). Clarification and standardization of substance abuse terminology. *Journal of the American Medical Association, 259,* 555–557.

Robinson, S. M., Sobell, L. C., Sobell, M. B., Arcidiacono, S., & Tzall, D. (2014). Alcohol and drug treatment outcome studies: New methodological review (2005–2010) and comparison with past reviews. *Addictive Behaviors, 39,* 39–47.

Roden, D. M., Xu, H., Denny, J. C., & Wilke, R. A. (2012). Electronic medical records as a tool in clinical pharmacology: Opportunities and challenges. *Clinical Pharmacology & Therapeutics, 91*(6).

Roizen, R. (2000). *E.M. Jellinek and all that!: A brief look back at the origins of post-repeal alcohol science in the United States*. Lecture prepared for the annual meeting of the ABMRF, San Francisco, CA, October 20–26, 2000. Available at: http://www.roizen.com/ron/jellinek-pres.htm

Samet, J. M. (2014). The Surgeon Generals' reports and respiratory diseases, from 1964 to 2014. *Annals of the American Thoracic Society, 11,* 141–148.

Samet, J. M. (2016). Epidemiology and the tobacco epidemic: How research on tobacco and health shaped epidemiology. *American Journal of Epidemiology, 183*(5), 394–402.

Scott, J. C., Woods, S. P., Matt, G. E., Meyer, R. A., Heaton, R. K., Atkinson, J. H., et al. (2007). Neurocognitive effects of methamphetamine: A critical review and meta-analysis. *Neuropsychology Review, 17,* 275–297.

Serre, F., Fatseas, M., Debrabant, R., Alexandre, J., Auriacombe, M., & Swendsen, J. (2012). Ecological momentary assessment in alcohol, tobacco, cannabis and opiate dependence: A comparison of feasibility and validity. *Drug and Alcohol Dependence, 126,* 118–123.

Sevigny, E. L., & Fuleihan, B. (2015). Measurement and design challenges in the study of drugs and society. In H. H. Brownstein (Ed.), *The handbook of drugs and society*. West Sussex, UK: Wiley.

Shiffman, S. (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment, 21,* 486–497.

Singh, T., Arrazola, R. A., Corey, C. G., Husten, C. G., Neff, L. J., Homa, D. M., et al. (2016). Tobacco use among middle and high school students—United States, 2011–2015. *Morbidity and Mortality Weekly Report, 65*, 361–367.

Sloboda, Z. (2012). The state of support for research on the epidemiology, prevention, and treatment of drug use and drug use disorders in the USA. *Substance Use & Misuse, 47,* 1557–1568.

Sloboda, Z. (2014). Reconceptualizing drug use prevention processes. *Adicciones, 26,* 3–9.

Sloboda, Z., Cicero, T., Inciardi, J., Leukefeld, C., McBride, D., & Stephens, R. (2009a). Introduction to the special issue. *Journal of Drug Issues, 39,* 1–4.

Sloboda, Z., Stephens, P., Pyakuryal, A., Teasdale, B., Stephens, R. C., Hawthorne, R. D., et al. (2009b). Implementation fidelity: The experience of the Adolescent Substance Abuse Prevention Study. *Health Education Research, 24,* 394–406.

Sloboda, Z., Stephens, R. C., & Alemagno, S. (1998). Postscript: Where do we go from here? *Journal of Psychoactive Drugs, 30,* 307–313.

Smith, S. M., Dart, R. C., Katz, N. P., Paillard, F., Adams, E. H., Comer, S. D., et al. (2013). Classification and definition of misuse, abuse, and related events in clinical trials: ACTTION systematic review and recommendations. *Pain, 154*, 2287–2296.

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., et al. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science, 14*, 319–351.

Suffoletto, B., Callaway, C., Kristan, J., Kraemer, K., & Clark, D. B. (2012). Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. *Alcoholism, Clinical and Experimental Research, 36,* 552–560.

Sung, E., Richter, L., Vaughan, R., Johnson, P. B., & Thom, B. (2005). Nonmedical use of prescription opioids among teenagers in the United States: Trends and correlates. *Journal of Adolescent Health, 37,* 44–51.

Sussman, S., Stacy, A. W., Johnson, C. A., Pentz, M. A., & Robertson, E. (2004). A transdisciplinary focus on drug abuse prevention: An introduction. *Substance Use & Misuse, 39,* 1441–1456.

Sussman, S., & Unger, J. B. (2004). A "drug abuse" theoretical integration: A transdisciplinary speculation. *Substance Use & Misuse, 39,* 2055–2069.

Thomas, Y. F., & Conway, K. P. (2010). The epidemiology of drug abuse: How NIDA stimulates research. In L. M. Scheier (Ed.), *Handbook of drug use etiology: Theory, methods, and empirical findings*. Washington, DC: American Psychological Association.

Treno, A. J., Marzell, M., Gruenewald, P. J., & Holder, H. (2014). A review of alcohol and other drug control policy research. *Journal of Studies on Alcohol and Drugs*, s17, 98–107.

Trujillo, K. A., Castaneda, E., Martinez, D., & Gonzalez, G. (2006). Biological research on drug abuse and addiction in Hispanics: Current status and future directions. *Drug and Alcohol Dependence, 84S,* S17–S28.

Turkkan, J. S., Kaufman, N. J., & Rimer, B. K. (2000). Transdisciplinary tobacco use research centers: A model collaboration between public and private sectors. *Nicotine & Tobacco Research, 2,* 9–13.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*(5365), 867–873.

Turner, C. F., & Miller, H. G. (1997). Monitoring trends in drug use: Strategies for the 21st century. *Substance Use & Misuse, 32,* 2093–2103.

Turner, C. F., Villarroel, M. A., Rogers, S. M., Eggleston, E., Ganapathi, L., Roman, A. M., et al. (2005). Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of telephone audio-CASI. *Addiction, 100*(10), 1432–1444.

Turner, L., Mermelstein, R., & Flay, B. (2004). Individual and contextual influences on adolescent smoking. *Annals of the New York Academy of Sciences, 1021,* 175–197.

Voas, R. B., & Fell, J. C. (2010). Preventing alcohol-related problems through health policy research. *Alcohol Research and Health, 33,* 18–28.

Volkow, N. D., & Baler, R. D. (2014). Addiction science: Uncovering neurobiological complexity. *Neuropharmacology, 76,* 235–249.

Warren, K. R., & Hewitt, B. G. (2010). NIAAA: Advancing alcohol research for 40 years. *Alcohol Research and Health, 33*(1 & 2), 5–17.

Weglarz, C. (1987). Rutgers Center of Alcohol Studies Library: A brief history. *British Journal of Addiction, 82,* 833–840.

Weisner, C., Schmidt, L., & Tam, T. (1995). Assessing bias in community-based prevalence estimates: Towards an unduplicated count of problem drinkers and drug users. *Addiction, 90*(3), 391–405.

Westermeyer, J. (1990). Methodological issues in the epidemiological study of alcohol-drug problems: Sources of confusion and misunderstanding. *American Journal of Drug and Alcohol Abuse, 16*(1–2), 47–55.

Wickson, F., Carew, A. L., & Russel, A. W. (2006). Transdisciplinary research: Characteristics, quandaries and quality. *Futures, 38,* 1046–1059.

Willenbring, M. L. (2010). The past and future of research on treatment of alcohol dependence. *Alcohol Research and Health, 33*(1 & 2), 55–63.

Wishart, D. S. (2005). Bioinformatics in drug development and assessment. *Drug Metabolism Reviews, 2,* 279–310.

Rockefeller Archive Center. (n.d.). http://www.rockarch.org/collections/rockorgs/bsh.php

Young, A. M., Glover, N., & Havens, J. R. (2012). Nonmedical use of prescription medications among adolescents in the United States: A systematic review. *Journal of Adolescent Health, 51,* 6–17.

Young, M. M., Dubeau, C., & Corazza, O. (2015). Detecting a signal in the noise: Monitoring the global spread of novel psychoactive substances using media and other open-source information. *Human Psychopharmacology, 30,* 319–326.

2 Transdisciplinary Research Perspective: Epidemiological Criminology as an Emerging Theoretical Framework for Substance Abuse Research

Jennifer M. Reingle Gonzalez and Timothy A. Akers

2.1 Transdisciplinary Thinking: Going Down the Rabbit Hole

In the polemic prose of Lewis Carroll's "Alice in Wonderland," Alice ventures down the rabbit hole, fraught with unknown peril, a chilling reminder that what we once knew or believed, can come to a screeching halt, drastically and dramatically, changing the very essence of our thoughts, thereby altering our once coveted place in the sciences. As scientists, generally, our space and place is a narrow world that should sit on the precipice, the edge of embracing new thoughts and new approaches in examining substance abuse research. Yet, our once strongly held world view can, at times, become more of an illusion, and come tumbling down, when mixed with other realities, processes, thoughts, methods, sciences, and perspectives; a tapestry of complexity. Such is the case when we go down the rabbit hole of science to eventually find that we must challenge, at times alone, the ethos of scientific traditionalism and silo thinking. Many of our brethren in substance abuse research fears charting new territory, new byways and pathways to discovery, new challenges that can lead to an amalgamation of transdisciplinary thoughts, ideas, and emerging methods and theories. This unfortunate reality avoids venturing out into the light of day for new and innovative scientific ideas.

In this chapter, we will argue that we must embrace a polymorphism of transdisciplinary thinking to better undertake the changing nature of substance abuse research, transcending both the researcher themselves and the training they are provided. We are oftentimes reminded that our staunch view of the world as discipline-specific scientists has conditioned us to examine the world with a narrowly refined lens, as though we are looking at a petri dish, where boundaries are clear, and depth is at a cellular level. Yet, we are, at this space and time of substance abuse research, encountering a new era, a new renaissance, where enlightenment reigns supreme, and a single perspective or discipline might do more harm than good.

In this case, substance use and misuse fits the paradigm of both a public health and criminal justice problem. Substance abuse and dependency costs taxpayers more than $534 billion in the United States each year (NIDA 2007), while drug use, possession, and distribution results in the incarceration of 14–19% of the total prison population, including both federal and state inmates, respectively (Mumola and Karberg 2006).

J.M. Reingle Gonzalez (✉)
Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas School of Public Health, 6011 Harry Hines Blvd., V8. 112, Dallas, TX 75390, USA
e-mail: Jennifer.reingle@utsouthwestern.edu

T.A. Akers (✉)
Division of Research and Economic Development, Morgan State University, 1700 E. Cold Spring Lane, Earl S. Richardson Library, Baltimore, MD 21251, USA
e-mail: timothy.akers@morgan.edu; drtimakers@gmail.com

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_2

However, when we examine how to address a transdisciplinary problem, such as substance abuse, conceptually and methodologically, we need to become aware of the unique distinctions between public health and criminal justice systems organizationally, theoretically, and methodologically. For example, although public health officials emphasize prevention of drug abuse and dependence, the need to consider redirecting existing or future prevention resources to better take into account a symbiotic relationship between public health and criminal justice may be more scientifically, fiscally, and logistically reasonable and efficacious. While some could argue that the criminal justice system is 'retributively-oriented' because of high drug-related recidivism rates, it should also be noted that their primary mission and role has been to prevent harm (through enforcement and sanctioning) employing various forms of prevention punishment for violating drug laws. The same argument can be directed to public health systems, particularly, in situations when protecting the public's health might outweigh the perceived benign nature of public health and its impact on substance abuse prevention (Akers et al. 2013). In other words, the current state of disjunction between the public health and criminal justice system results in a surplus of untreated adolescent and adult repeat offenders who may be amenable to rehabilitative treatment that shares in the resources and expertise of both systems of prevention and enforcement.

2.2 Rethinking Health and Crime in Substance Abuse Research

Substance abuse research, a subject fraught with diverse perspectives and discipline-specific peril, has been studied from many disciplinary perspectives, ranging from psychology to economics, each with their own seemingly unique and unyielding insight. While commendable, each discipline brings its brand of science, modus operandi, and few models that are transdisciplinary enough to breakdown the scientific walls of isolationism. The same can be said for the science and practice of public health and criminal justice. Take public health, for example. Public health researchers have been employing a plethora of interventions and treatment, evaluating policies, and assessing the etiology of substance abuse for decades. At the same time, criminal justice researchers have created and evaluated drug courts, treatment programs in correctional settings, and policing tactics intended to reduce drug abuse and dependence. For decades, criminal justice practitioners and scholars have even approached primary, secondary, and tertiary prevention strategies for youth to try and change risky behavior before, during, and after it can take deep root and start to cultivate a growing milieu of at-risk behavior (Brantingham and Faust 1976). Despite the clear overlap in the public health and criminological research agendas, research methods, and itinerant theories, there has been little discussion between researchers in these two disciplines (Akers and Lanier 2009). Unfortunately, transdisciplinary communication in drug abuse research has virtually gone almost unnoticed, and certainly has not been rewarded or encouraged. In fact, scientific colleagues and senior administrators in higher education who are wedded to perpetuating their discipline-specific brand of theory have, either unintentionally or intentionally, discouraged this cooperation between disciplines. Invariably, when you look at a scientific article and the disciplines of each author, they tend to be the same: psychologists publish with psychologists, sociologists with sociologists, and public health scholars with public health scholars; this point becomes clear.

However, when we examine substance abuse and the potential myriad of transdisciplinary theories and methods that can be conceptualized, developed, and integrated, it becomes an excellent research domain for scholars and practitioners of criminal justice and public health to stretch the methodological limits to understand drug use behavior (Potter and Rosky 2013). For example, irrespective of whichever substance abuse methods are employed, despite having clear physiological health effects (for a review, see Boles and Miotto 2003), the use of some

drugs is societally condoned (e.g., alcohol, tobacco, and, in many cases, marijuana), while others are deemed illicit and societally unacceptable. As a result, a unique dichotomy emerges whereby problematic use of licit, legal substances is rectified via treatment services, while problematic use of illicit, illegal substances is often punishable through jail time or other retributive and punitive measures.

This social construction of the harm, or, more accurately, the physiological risk associated with each type of substance is particularly interesting in that some illicit drugs have been touted as being less harmful than alcohol or tobacco (Macleod et al. 2004). For instance, alcohol and tobacco have clear negative health effects that have been long established in the scientific literature, including multiple types of cancer and organ damage. Marijuana, on the other hand, has been linked with minor respiratory disorders, but few long-term poor health outcomes (NIDA 2007). However, the natural experiments in decriminalization of marijuana, legalization, and healthcare record uniformity, which are currently occurring in the United States, provide an optimal research environment for scientists to learn about the long-term outcomes of marijuana use in comparison to alcohol and tobacco. In addition, some illicit use is condoned in some jurisdictions (for instance, marijuana has been decriminalized for medical usage in the State of California, but remains a Schedule 1 drug federally in the United States). Therefore, given the contradictory legislation of the current date (2014), the study of marijuana is an especially complex methodological process.

To provide further detail on the complexity associated with the study of illicit (or in some states, decriminalized or legalized) substances, Colorado (among other states) has recently legalized the open and legal use, sale, distribution, and manufacturing of marijuana. Therefore, in the state of Colorado, marijuana is no longer being designated as a controlled substance. Over time, data on substances, such as marijuana, will inevitably become more available through new venues of reporting and data collection methods for clinical outcomes about marijuana's effects and other health outcomes, simply because it will now be more freely and legally reported to healthcare providers and law enforcement. In other words, given that research on drug use and abuse is particularly complex in light of the social environment and new policy changes surrounding drug use in the United States, the states that are opting to legalize or decriminalize formerly illicit substances (such as marijuana) will, for better or worse, lead to a richer and more robust storehouse of surveillance data. In theory, substance abuse researchers will, ideally, transcend disciplinary boundaries due to the freely available, medically verifiable outcome data (for instance, health records, tax information, sales and quantity distributed, and police/legal data) beyond the traditional self-reported information that is used today. And, as the distinction between drugs as pharmaceuticals and drugs as criminal behavior become increasingly blurred, the need for integrated surveillance systems and protocols for substance abuse research will become even more paramount (Akers 2013).

2.2.1 Language and Lexicon: Finding a Common Ground

Language, that primordial stew of letters, words, numbers, characters, or gestures, pours out its messages and meanings across countless dialects, mathematics, and scientific terminology. Substance abuse research is no different. Across endless ocean of terms used to convey scientific meaning, substance abuse researchers bring forth their own unique brand, discipline, and blending of terms; because what is a science if it was not for its own terminology? The world of substance abuse research calls forth many terms. The sheer complexity is reflected in the disparity in terminology that masks a distinction between the terms 'substance abuse' and 'drug abuse.' Take a moment to consider how seemingly similar terms grew and replaced other terms for decades. The

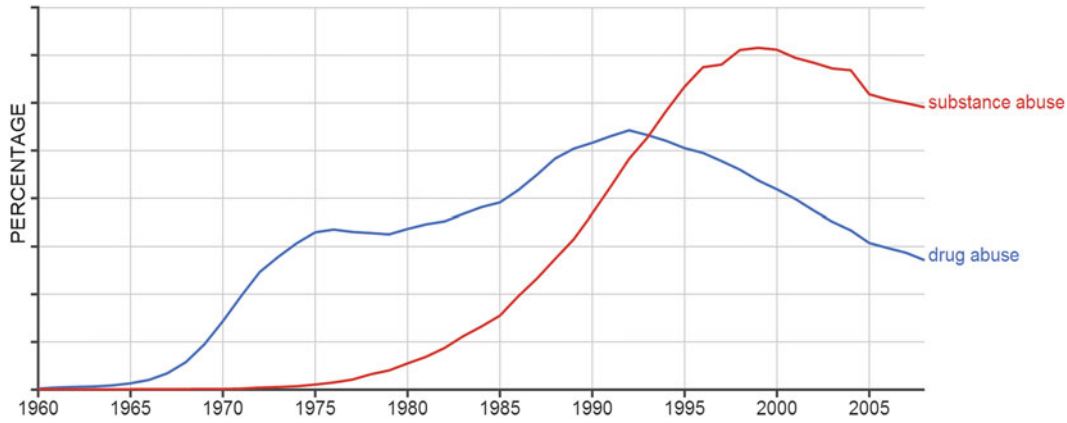

Fig. 2.1 Substance abuse and drug abuse Google Ngram terms used from 1960 to 2008

'Google Ngram' provides a visual image of word usage in books over the decades (see https://books.google.com/ngrams). This sophisticated program extricates words used in tens of millions of books authored in English over the generations and provides a graphic depiction of a terms usage over time and percent, though the percent indication is not relevant as compared to the portrayal of the longitudinal image.

To illustrate our point, we chose the description of 'substance abuse' and 'drug abuse' and juxtaposed these similar terms over the decades. Figure 2.1 illustrates the ebb and flow in the use of the terms, which have changed over the decades, thereby, possibly, altering our scientific outlooks as to how best to synthesize or use these terms across different disciplines, or even how possibly to tailor, design, or weave new interventions. While this distinction can most certainly encompass volumes of scholarly words, ours is not to debate this point, but rather, to illustrate how language, descriptions and designations can change the very fabric of our weave. One may think this is an innocuous distinction, but we appreciate a quip made by the late George Orwell in his famous *1984*, where he scripts that "*But if thought corrupts language, language can also corrupt thought*." From our perspective, transdisciplinary thinking might serve to avoid corrupting both language and thought.

2.3 Drug Users (a Challenging Group to Research): Defining a Research Protocol

Over the course of three decades (1970–2006), drug abuse has soared 354.7%, from 415,600 to 1,889,810 cases, respectively, in the United States (Benson 2009; U.S. Bureau of Justice Statistics 2011). Identifying drug users, as mentioned above, presents a unique problem for drug abuse research. Who are drug 'users,' 'misusers,' and 'abusers'? Is using marijuana once a month 'misuse'? At what point does 'misuse' become 'abuse'? Public health and criminal justice researchers approach all of these questions differently as a function of the highly specific research training they received. For example, "in 2010, an estimated 22.1 million persons (8.7% of the population aged 12 or older) were classified with substance dependence or abuse in the past year based on criteria specified in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV-TR). Of these, 2.9 million were classified with dependence or abuse of both alcohol and illicit drugs, 4.2 million had dependence or abuse of illicit drugs but not alcohol, and 15.0 million had dependence or abuse of alcohol but not illicit drugs" (CBHSQ/SAMHSA 2013).

In this example, the American Psychological Association (2013) clearly defines substance use disorders in terms of 11 specific criteria: (1) taking larger amounts of the substance than originally intended; (2) wanting to cut down on use but unable to do so; (3) drug seeking; (4) recurrent legal problems; (5) unable to conduct routine activities at home, work, or school due to drug use; (6) continued use despite problems associated with the substance; (7) giving up recreational activities as a result of drug use; (8) continued use despite harms; (9) tolerance; (10) withdrawal; and (11) continued use despite physical or psychological problems associated with drug use. The number of criteria met was then categorized into 'abuse' or 'dependence' to classify each individual. These criteria were developed to broadly apply across all types of drugs, from alcohol to methamphetamines. The status of licit versus illicit was not a factor in terms of the diagnosis.

As further evidence of intra-disciplinary complexity in measuring drug use, in 2014, with the release of the DSM-V, the APA itself has modified how substance use is classified (removing legal problems and adding 'strong desire or urge to use a substance' as an indicator of substance use disorders). The more criteria met for a diagnosis, the more severe the disorder, and the terms 'abuse' and 'dependence' are no longer used to refer to any substance. Instead, all criteria are weighted equally on a continuum, and any person who indicates 2 or 3 of the 11 criteria are considered to have a 'mild substance use disorder,' those who report 4 or 5 have a 'moderate disorder,' and those with 6 or more of the 11 criteria are categorized as having a 'severe substance use disorder' (Grohol 2013). Although there are pros and cons associated with each variety of measurement, the APA provides an interesting example of how the definitions and categorizations of substance use, abuse and dependence change within one discipline over a very short amount of time.

None of these APA definitions are exhaustive; in fact, a large proportion of the literature on substance abuse does not operationalize drug abuse using these definitions. For instance, a great deal of substance abuse research in criminal justice and criminology uses official reports derived from police records. In this case, drug use will likely be defined in terms of a drug-related arrest (possession of drugs and/or paraphernalia, possession with the intent to distribute). Conversely, some public health studies have used emergency rooms as venues to recruit substance abusers that present for treatment of an overdose, violence, or drug-related accident. In effect, these sampling methods identify some of the most at-risk substance users; however, these individuals do not necessarily meet the definition used by the APA. That is, the cases may or may not be drug abusers per se; instead, they may be occasional users or involved in distribution of illicit substances, thereby leading to what we might coin as a methodological or conceptual 'discordance.' That is, as disciplinary lines become blurred (as you can clearly see in the example above), research methods begin to overlap.

As methods continue to overlap, substance abuse researchers will further cross over and morphous into other realms of science, and a new enlightenment, a new renaissance for the science of substance abuse research will emerge. Although we are not promoting one discipline, method of sampling, or even one theory or practice over another, we are directly, encouraging researchers to reassess their target populations. This in-depth thinking to determine the most valid scientific method, rather than the most convenient one, will help determine which method of selection, sampling, measurement, theoretical modeling, and analysis is most appropriate given their research question. In many cases, the optimal sample, or the methodological discordance, may require more transdisciplinary thinking about the most meaningful definition of substance use, abuse, and dependence, is of utmost importance for each specific study. Although quantity of research produced is often rewarded over quality of science and the impact of a publication on practice, we urge those who have the power to encourage ***critical thinking*** in their operationalization of substance use, abuse, and dependence in their departments

and research circles. We recognize that this thinking is far more time consuming than simply using what has been used in 'previous research,' the potential for progress is tremendous if this type of culture can be developed.

2.4 Methods at the Intersection of Epidemiology and Criminology

Tapping into the collective wisdom of transdisciplinary thought takes some creative thinking. One does not simply claim to be a winner when a race is only creeping along. Beyond these metaphors, we recognize that substance abuse researchers, first and foremost, are a tapestry of diversity with respect to their disciplinary lineages, thought patterns, and many diverse (or similar) approaches used to study substance abuse. Their training may range from biology to social work and any and every discipline in between. Yet, there tends to be a plethora of methods used to study substance abuse, and these methods are not rooted in any one discipline; though, arguably, they tend to employ more epidemiological measures with respect to their reporting of health impact, pathology, disease transmission, morbidity and mortality outcomes.

Historically, at the intersection between epidemiology and criminology lays similarity in common methods used. For example, criminological research was, for many years, predominantly cross-sectional in nature, providing only a single snapshot of criminal and deviant behavior. In the early 1990s, criminologists heatedly debated the value of longitudinal designs in their research. Eventually, a consensus emerged that longitudinal research designs were superior to cross-sectional designs, in that the findings were more valid and replicable (Menard and Elliott 1990). Books were published re-testing theories, such as developmental life-course theories of adolescent behavior, with longitudinal data as its scientific grounding and method of choice (Liberman 2008). In support of this approach, the field of criminal justice has embraced the notion that longitudinal data are valuable sources of information, despite the tremendous cost, labor, and time associated with data collection.

Through the evolution of greater transdisciplinary thinking, longitudinal designs in epidemiology and criminology have allowed both disciplines to expand into new directions and embrace new horizons, while, at the same time, working more closely together, and sharing theories, methods, and hypotheses (Akers et al. 2013). In criminology, longitudinal data have been used to describe intra-individual trends in arrests and drug use over time. Similarly, social epidemiologists used longitudinal data to study the long-term effects of poverty, sugar-sweetened beverages, or limited access to healthy foods (Drewnowski and Specter 2004; Kendzor et al. 2012). These problems could not be studied with cross-sectional data, as the single time point of data collection limits a researchers' ability to evaluate change over time. Longitudinal data are especially appropriate in both epidemiology and criminology, as behavior (particularly substance use behavior) changes measurably over time.

As we continue to keep focused on substance abuse and its many complexities from a research methods perspective, we note that cross-sectional and longitudinal designs are not the only research designs used to study substance abuse problems, although they are the most common. Case-control studies are often employed in epidemiological studies, particularly, in the case of rare outcomes (e.g., anthrax poisoning or Hepatitis C). These designs are also useful in time-sensitive situations, such as outbreak investigations used in epidemiology to detect and control the source of an infection (for instance salmonella or listeria in a hospital). Briefly, a researcher would identify targeted cases (those who have listeria) and compare to a series of similar, but non-diseased (often matched) controls. Once the sample has been identified, the researcher will look backwards in time to identify which risk factors (or protective factors) differentiate the cases and controls. In other words, *which variables increase the risk of disease?* We would find that those patients with listeria were operated on in the same operating room and the surgical tools were not cleaned properly. This

will help us in preventing future cases of listeria. In substance abuse research, we might consider case-control studies as 'mining the discipline.' That is, we can pick away at all the factors, all the minutia, that might have contributed to a person or event being subjected to a known or unknown risk. In other words, we can look at all kinds of variables that might increase or decrease the risk of abusing substances from multiple disciplines, including neurology, sociology, demography, epidemiology, and criminology, among many others. It is critical to note that a case-control method transcends disciplinary boundaries, but may not necessarily use the same lexicon, as described in a criminology or criminal justice methods book or encyclopedia (Piquero and Piquero 2002). No discipline can hold claim to this method per se. Although these research designs have been pervasive in medicine and public health, they are just now being used to assess drug use and crime. In fact, most of the criminal justice research using case-control designs is published in transdisciplinary journals across the fields of criminal justice, public health, psychiatry, and medicine (Needleman et al. 2002; Reingle et al. 2013). Once transdisciplinary communication increases, these types of designs will be particularly useful in identifying factors that developmentally contribute to initiation, continuity, and desistance from various substances.

2.5 Rethinking Our Way of Thinking for Substance Abuse Research

The design of studies in criminal justice and public health is not the only thing that has been changing over time. New analytical methods have paved the way for hierarchical analysis that can occur at the ecological level (macro), meso, exo, and micro-level (Akers et al. 2013; Trickett and Beehler 2013; Akers and Lanier 2009). These methods allow for new ways of thinking about old problems, as each layer (e.g., ourselves, our family, our friends, our community, and our society, for instance) has an effect on our behavior, and we have an effect on theirs. For example, the 'micro' level unit of analysis includes characteristics of the individual, including the home environment, religious institutions, and the workplace. In the micro-environment, people interact directly with their environment. The meso-environment connects micro-environments, or interactions between people (e.g., teachers and parents, child and parents, etc.), and the exo-environment has no direct (only indirect) influences on the individual (including communities or neighborhoods to which a person belongs). And the 'macro' unit of analysis refers to the environment in which we live, work, and play, such as the culture of the United States, or our particular state of residence. Changes in policy, such as the Affordable Care Act or federal sentencing guidelines, occur at the macro level.

These various layers, dimensions, or units of analyses can serve as sort of ecological links in a chain that helps to frame transdisciplinary thinking. Take, for instance, impaired driving as a concrete example of how ecological information is used in an epidemiologic and criminological study. Impaired driving (or driving under the influence of alcohol) is a problem that spans both the public health and criminal justice research agendas. In other words, impaired driving has often been treated at the individual (micro) level by criminologists using police arrest reports. Recently, epidemiologists have employed 'big data' (macro-level, integrated database systems) on motor vehicle crashes and fatal accident reports to assess the influence of alcohol-related policies on impaired driving. Integrative analytic methods now allow researchers to determine if the launch of a new policy has any effect on deaths or injury as a result from impaired driving. Interrupted time-series analysis is one such technique that can help to understand micro-, meso-, or exo-system level events over time to determine if there was a spike across public health communication or criminal justice enforcement type of macro-level policies. This method will be especially appropriate to evaluate changes in health outcomes as marijuana legalization policies change over time.

In summary, transdisciplinary thinking about driving under the influence of alcohol has now resulted in several policy changes (e.g., bartender training, driver's license suspension programs, and mandatory jail sentences) that have the potential to reduce the number of accidents and deaths attributable to impaired driving. To address the exo- and meso- layers, many substance abuse prevention and intervention programs now include families, peers, and teachers. It is clear that rehabilitation cannot be successful in a vacuum, as people will return to their daily lives and, without grounding in the proper skills set, will revert to continuing to use alcohol or drugs.

The vast majority of research today occurs at the micro-level, collecting survey data (or crime records or hospital data), on each individual person. The unit of analysis is the patient (or offender). As a result of these micro-level designs, the implications of this research apply to individuals. The same example can be applied to an HIV/AIDS positive person who intentionally infects sex partners without their knowledge, or intentionally shares their contaminated injection drug works to unsuspecting prey. Yet, from a transdisciplinary perspective, we must increasingly consider how our analyses and units of analyses overlap or intersect in order to enrich our analytical thinking in the area of substance abuse research. As noted above, a transformation in thinking about our unit of analysis, one that transcends disciplinary boundaries, to take into account all levels, such as the macro-level that has the potential to create far more change (via public policy), as the results of a well-designed macro-level transdisciplinary study could be generalized to the entire country. Policy evaluations conducted at the exo- or macro-levels have the potential to decrease risk behavior (e.g., impaired driving), or increase positive behavior (e.g., use clean needles, wearing of condoms, or undergoing routine drug screening), for everyone in the macro environment. These types of designs that apply to large portions of the population are likely to have the greatest impact on the public health, or substance abuse research.

2.6 Challenges in Conducting Transdisciplinary Research on Substance Abuse

2.6.1 Traditionalism Versus Enlightenment

Apart from what has already been shared, as substance abuse scientists, practitioners, policymakers, or clinicians, we share a kind of cognitive dissonance, a 'love-hate' relationship battling inside our colleagues and ourselves when it comes to embracing or letting go of traditionalism versus enlightenment. On the one hand, it is difficult to let go of preordained methods, approaches, techniques, and models that we were trained in extensively and have used for decades; while, at the same time, trying to embrace a new era of enlightenment, where greater numbers of diverse disciplines are drawing from the science and experience of others. To address these issues, we feel it is necessary to take baby steps to share what has led us to strongly embrace a merger of epidemiologic and criminogenic integration within the context of substance abuse research.

To begin with, it is important to note that we are not simply suggesting that we replace one discipline for another, or to go down one rabbit hole rather than the other. Quite to the contrary, what we are suggesting is that we as researchers consider the possibility of integrating disciplines; particularly, epidemiology and criminology, which may serve as a channel to bring forth new and innovative trans-disciplinary thinking, while holding on to the rich experience and history of other disciplinary streams of consciousness, training, experience, and insight. Take, for example, psychology, sociology, social work, criminology, and epidemiology (among others). Each of these disciplines coronets a majesty and tapestry of expertise and diversity of thought. Yet, each also brings with it, for better or worse, a myopic disciplinary way of thinking; a one size fits all sort of mindset. That is, psychologists will look at the individual; sociologists will examine the group; social workers the

case services; epidemiologists the data; and criminologists, the aberrancy of the individual and group dynamics. Clearly, given the complexity of the substance use as a societal problem, all of these disciplines has a unique insight to contribute that may serve to help us learn and by extension, prevent and reduce the costly burden of substance use and abuse in the United States.

2.6.2 Bias While Tiptoeing Through the Tulips (Sample Selection and Reporting of Drug Use)

All disciplines, regardless of heritage or legacy, seek to understand the challenges in conducting substance abuse research from their single disciplinary perspective. Through our myopic lenses, we draw conclusions, make assumptions, and report findings (Whoriskey 2012). We are, for all intents and purposes, biasing our research, biasing our samples, and biasing our findings, when we neglect to take into account unique and diverse perspective on a singular problem. This does not mean that researchers *must* accept alternative views; however, we are asserting that substance abuse researchers simply consider the arguments and potential alignments from disciplines other than their own. An appropriate analogy might be how one views their own children, as sometimes our judgment is clouded because we are so close to the subject (in this case, our children) that our own perceptions distort our objectivity or our ability to view the behavior of our children from a single person's perspective.

From the lowly researcher to a vast research group, the world and science of substance abuse research is plagued by bias, innuendo, and silo mentality and methodology. While such a statement will, most certainly, create controversy and most likely invoke animus, the intent is, first and foremost, to stimulate debate and dialogue. Attacking research biases requires discussion and collaboration across disciplinary domains. However, a research cohort that has or only embraces scientists trained in the same discipline is not a science any longer; rather, it is an advocacy group. Such is the need to call for new perspectives, by identifying where disciplines share a common core, a common language, and a common understanding, but are different and unique nonetheless.

2.6.3 Case Study: The Evolution of Epidemiological Criminology

To avoid extinction or prey, species throughout the world have had to constantly adapt to their environments. From the deepest recesses of the oceans to the highest mountain peaks, the animal kingdom has learned to accept change as a normal course of their evolution, to find a middle ground that provides the greatest amount of opportunity. Not unlike the evolution of species, the scientific community has learned to embrace its subject of study, in this case 'substance abuse' by encouraging a vast array of disciplines to bring their plethora of theories, concepts, methods, and practices to the science of substance abuse research. Yet, rarely does a new discipline evolve to transcend a diversity of scientific domains while, at the same time, embracing the richness unique to each discipline.

In the recent past, a breakthrough in transdisciplinary thinking has emerged in the development of the *Epidemiological Criminology* paradigm (Akers et al. 2013; Akers and Lanier 2009). As you may have gathered from reading this article, the research agendas (particularly in regards to substance abuse) in criminal justice, criminology, public health, and epidemiology are largely complementary and overlapping. This integrationist trend towards a transdisciplinary ethos has been evidenced by the U.S. Centers for Disease Control and Prevention, the National Institute of Justice, and the National Institutes of Health proclamation that violence has now been identified as a public health imperative and national priority research area. Specifically, since the latter decade of the 1980s and continuing to the early 1990s, the world of public health has

targeted new prey (e.g., violence and violence-related injury prevention) in their quest to encourage transdisciplinary research. Arguably, this awareness may have been driven by a number of U.S. surgeon generals who publicly proclaimed violence as a public health imperative (Satcher 1995; Sullivan 1991; Koop 1989).

Crime and violence, as well as substance abuse, have been studied across a plethora of disciplines, including nursing and occupational safety (workplace violence and impairment on the job), environmental safety (design of safe neighborhoods, disposal of drug injection equipment), and public health (prevention of violence and drug use) (Krug et al. 2002). However, within the context of crime, violence, and substance abuse research, the fields of criminology and criminal justice have long served as primary disciplines in addressing these areas, having their anchor origins in the science of sociology. Yet, both public health and criminal justice researchers continue to travel down parallel pathways in their shared quest to reduce and understand drug use, crime, and violence, while continuing to hold steadfast to their uniquely rich disciplines.

For instance, public health officials and researchers are consistently emphasizing the need for primary prevention of violence, drug use, and other risk behavior. That is, although primary prevention (or eliminating the onset of drug use or violent acts) is addressed in criminal justice research, it can be more challenging to implement these prevention programs due to the reactive nature of the criminal justice setting (whomever encounters the criminal justice system, by definition, are no longer a candidate for primary prevention, as they have already initiated drug use and/or crime and have made contact with the police). To illustrate further, the police are the first point of contact when someone enters the criminal justice system. By the time a person comes in contact with the police, it is typically because drug use (or any other criminal behavior) has already initiated.

It is not impossible to conduct primary prevention of crime or drug use using a criminological framework, as innovative criminal justice programming has integrated school-based primary prevention of drug use and violent crime (Webster-Stratton and Taylor 2001). However, these types of researchers are forced to transcend disciplinary boundaries to conduct this research. In addition, policing scholars are pushing to develop models of policing that will *prevent or reduce* crime, drug use, and violence (sometimes referred to as targeting the 'root causes,' such as through programs as McGruff the crime dog, or scared straight programs); however, the reactive nature of the policing institution still tend to present a challenge in the success of these efforts. Public health researchers, on the other hand, are often implementing prevention programs in schools and in community settings though they, at times, are developing and implementing interventions with little or no evidence to support their strategy. These two groups aim to achieve the same goal, drug use and crime prevention, but through different means. Part of the challenge of transdisciplinary substance abuse research is to identify a common thread whereby various disciplines can espouse expertise.

The *Epidemiological Criminology* paradigm recognizes that for a crime or delinquent act to have been committed, a statute or ordinance must have been violated. If no statute was violated in any way, no crime has occurred (Akers et al. 2013, Potter and Akers 2013). The operationalization (or definition) of 'legal' also varies over time. In Colorado, for example, the state has recently legalized the sale, distribution, and manufacturing of marijuana. What was once a crime in the state statute, such as the possession, sale, and distribution of marijuana, is no longer illegal. Hence, criminal justice interventions of interdiction or prevention may no longer apply. Whereas, on the other hand, public health interventions may continue to apply even more so, as in the case of adapting a similar approach to using a smoking cessation program or clean needle campaigns, especially from a transdisciplinary perspective for substance abuse research scientists (Nash et al. 2003).

The second level of prevention beyond primary is known as *secondary* prevention, or sometimes referred to as opportunity reduction or

risk reduction. This second tier intervention targets those who have already developed risk factors for the condition (for instance, an adolescent starts associating with a group of friends who use drugs, a clear risk factor for drug use), but the condition is not yet apparent in the individual (e.g., the individual does not use drugs themselves yet). To provide another example, the target audience for secondary prevention may be new smokers, or those who are surrounded by peers who use illegal drugs in an after-school program. These programs are more expensive than primary prevention, and are generally less successful once a risk behavior has been initiated. In this case, public health officials would seek to enroll at-risk candidates into treatment programs to reduce or discourage smoking and drug use, while police would identify these individuals (who may not be abusing or dependent upon substances) and seek incarceration, probation, or community service as a punishment intervention. However, given the especially high rates of recidivism among those released from penal incarceration, several criminal justice agencies are attempting new methods for reducing the number of offenders incarcerated (a New York State report found that a pilot test of drug treatment in lieu of incarceration was met with great 'success' and should be rolled out full-scale) (New York State Commission on Drugs and the Courts 2000). These approaches broaden tertiary prevention, whereby the perpetrator of an act has already carried out the violation (e.g., a person has used drugs), this approach is attempting to intervene before the person is incarcerated or subject to other legal and social ramifications. For example, more recently, drug courts and probationary treatment programs have emerged in the judicial system to encourage treatment for drug users and rehabilitation rather than incarceration. Although this type of thinking has been historically deemed 'political suicide,' taxpayers have become more invested in the idea of early intervention and treatment rather than dealing with the high costs of incarceration for decades.

Finally, tertiary prevention is a final stage attempt to treat a problem; in our case, substance abuse or dependence. This stage is the most expensive; as treatment must be intensive once physiological dependence has occurred. In many cases, individuals who self-identify as having a drug abuse problem will enroll in in-patient treatment. This stage focuses on the prevention of this behavior from spreading to or contaminating others, resulting in their substance abuse. Tertiary prevention is where containment occurs. From a transdisciplinary perspective, the paradigm of epidemiological criminology sets forth a unique framework that can help guide in the (re) design of a substance abuse research model that captures the most salient concepts and behaviors. For example, the criminal justice system, by and large, deals with drug dependent persons in the same way they deal with occasional users. The justice system, largely due to logistical complications, provides little differentiation of offenders (with the exception of inmate segregation to protect other inmates from especially violent individuals). However, if a particular court system is innovative, an offender may be assigned to a drug court or treatment program in lieu of incarceration (Hepburn 2005). This is where a tipping point between healthy and criminal behavior can sway an analysis from only one disciplinary perspective. Unfortunately, incapacitation does not address the root of the problem (the drug abuse), which frequently results in drug recidivism in comparison to drug courts or treatment-oriented sentences (Wilson et al. 2006).

2.6.4 Healthy Behavior or Criminal Behavior: Identifying a Tipping Point

Figure 2.2 depicts the *Epidemiological Criminology* model, which can help to conceptually frame a transdisciplinary approach to substance abuse research (see Akers et al. 2013). Theoretically and practically, a substance abuse researcher, regardless of their area of training,

should try and determine whether the substance abuse behavior tips more towards healthy (and experimental in nature) or criminal (see Fig. 2.2). This perspective provides the transdisciplinary researcher with a decision, a tool, on how best to begin their analysis. Built within the model are four determinants to help identify the myriad of life-course events that can tip behavior, thus activating the bio-psycho-socio- and environmental igniters that might have served as the catalyst to spark an aberrant and disparate trend of behavior. These factors can help in determining whether a criminal or deviant behavior is being, or has been, influenced by either internal or external, behavioral or biomedical disparities, that cultivate and nurture criminogenic outcomes. Early on in this chapter, we discussed briefly the differences between micro-, meso-, and macro-level analysis of behavior and environment. Each of these levels of influence, as well as the extent of their influence, should be considered. For instance, it is important to gather information as whether the drug use behavior is influenced by the individual themselves, their peers or associates, or encouraged by policies or the larger community in which an individual resides. For example, a change in policy can criminalize or decriminalize substance abuse violation (e.g., consider Colorado, as an illustration). Only time will tell whether the legalization of marijuana in Colorado will escalate the number of users who go into treatment or experience some facet of the public health or criminal justice system as criminally labeled substance abusers. To summarize, what is often viewed as enforcement by the criminal justice system, may also be enforcement from a public health perspective.

2.7 Summary

As we conclude with this transdisciplinary analysis for assessing new and emerging perspectives in substance abuse research methods, the introduction and evolution of *epidemiological criminology* as an innovative, transdisciplinary science has helped to enhance our understanding and express our need for scientific inclusion and scientific innovation. The idea that all social behavior, be it healthy or criminogenic, is intertwined and interwoven across disciplines is nothing new per se. To prevent or dissuade aberrant or substance abuse behavior, some theorists might argue that to cultivate and implant a healthy behavior and lifestyle a community should increase the amount of street lighting in a neighborhood thereby making a target less 'suitable' for transient behavior ripe for keeping substance abuse (*or substance abuse research enlightenment*) in the shadows, both literally and figuratively. Others disciplines and theories may also call for primary preventive measures which might include such street activities as instituting community watch, or have community meetings to strengthen bonds and social capital within the neighborhood. Although both of these options, practices, or theories might play a role in crime or substance abuse reduction within a community, these interventions may also have unintended, albeit positive, effects and outcomes, such as increased walking in the neighborhood (more physical activity) and social events, thereby creating a healthy mind, body, spirit and community.

The inverse, on the other hand, may play into the subconscious stressors associated with living in a previously disorganized neighborhood that is in the early stages of recovery. When the pendulum of power swings between healthy or criminal in which it meets that tipping point threshold at either end, having chosen a wrong intervention or theory can drastically alter behavior, community dynamics, or research findings for decades, be they positive or negative. This is why it is critical that we urge the reader to think 'outside of the box' and embrace a new renaissance, a new era of scientific openness when considering how substance abuse research and its myriad of aberrant, deviant, and criminal behaviors might be directly and indirectly woven together when they embrace healthy or unhealthy choices. Researchers and scholars have reiterated for decades that no single discipline should operate in a vacuum. However, espousing such a noble goal but failing to practice what one preaches are two separate things. Arguably, the

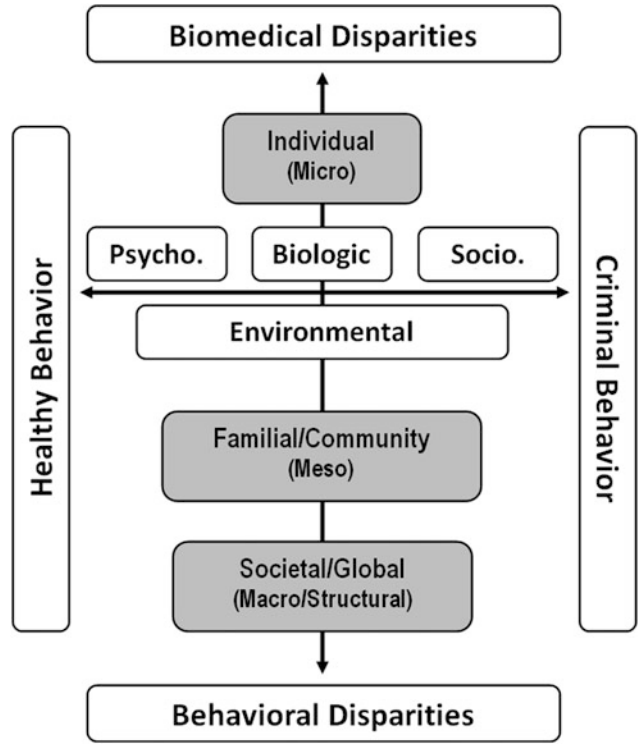

Fig. 2.2 An *epidemiological criminology* model for explaining criminal and deviant behavior

resources and infrastructure for transdisciplinary research, with a sway towards substance abuse research methodologies and theories, are already beginning to take shape and see a new light of day, a new enlightenment. Now, it is up to us as scholars and practitioners to push the boundaries, or venture down Alice's rabbit hole, to embrace the unknown and confront our fears in order to make a difference in the intellectual landscape of the future.

References

Akers, T. A. (2013). Criminological epidemiology or epidemiological criminology: Integrating national surveillance systems. In E. Waltermaurer & T. Akers (Eds.), *Epidemiological criminology: Theory to practice*. Routledge/Taylor and Francis: United Kingdom.

Akers, T. A., Potter, R. H., & Hill, C. (2013). *Epidemiological criminology: A public health approach to crime and violence*. San Francisco, CA: Jossey-Bass/Wiley and Sons.

Akers, T. A., & Lanier, M. M. (2009). "Epidemiological criminology": Coming full circle. *American Journal of Public Health, 99*(3), 397–402.

Benson, B. L. (2009). Escalating the war on drugs: Causes and unintended consequences. *Stanford Law and Policy, 20,* 293–357.

Boles, S. M., & Miotto, K. (2003). Substance abuse and violence: A review of the literature. *Aggression and Violent Behavior, 8*(2), 155–174.

Brantingham, P. J., & Faust, F. L. (1976). A conceptual model of crime prevention. *Crime and Delinquency, 22,* 284–296.

Center for Behavioral Health Statistics and Quality. (2013). *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings* (NSDUH Series H-41, HHS Publication No. (SMA) 11-4658). Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at http://www.oas.samhsa.gov/NSDUH/2k10NSDUH/2k10Results.htm

Drewnowski, A., & Specter, S. E. (2004). Poverty and obesity: The role of energy density and energy costs. *The American Journal of Clinical Nutrition, 79*(1), 6–16.

Grohol, J. (2013). DSM-5 changes: Addiction, substance-related disorders and alcoholism. *Psych Central*. Retrieved on February 24, 2014, from

http://pro.psychcentral.com/2013/dsm-5-changes-addiction-substance-related-disorders-alcoholism/004370.html

Hepburn, J. R. (2005). Recidivism among drug offenders following exposure to treatment. *Criminal Justice Policy Review, 16*(2), 237–259.

Kendzor, D. E., Caughy, M. O., & Owen, M. T. (2012). Family income trajectory during childhood is associated with adiposity in adolescence: a latent class growth analysis. *BMC Public Health, 12*(1), 611.

Koop, C. E. (1989). Introduction to injury prevention: Meeting the challenge. *American Journal of Preventive Medicine*, *5*(3) Suppl.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The lancet, 360,* 1083–1088.

Liberman, A. M. (Ed.). (2008). *The long view of crime: A synthesis of longitudinal research*. New York: Springer.

Macleod, J., Oakes, R., Copello, A., Crome, I., Egger, M., Hickman, M., et al. (2004). Psychological and social sequelae of cannabis and other illicit drug use by young people: A systematic review of longitudinal, general population studies. *The Lancet, 363,* 1579–1588.

Menard, S., & Elliott, D. S. (1990). Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency. *Justice Quarterly, 7* (1), 11–55.

Mumola, C. J., & Karberg, J. C. (2006, October). *Drug use and dependence, state and federal prisoners, 2004*. Bureau of Justice Statistics Special Report (NCJ 213530). Available online at http://www.bjs.gov/content/pub/pdf/dudsfp04.pdf

Nash, J. M., Collins, B. N., Loughlin, S. E., Solbrig, M., Harvey, R., Krishnan-Sarin, S., et al. (2003). Training the transdisciplinary scientist: A general framework applied to tobacco use behavior. *Nicotine & Tobacco Research, 5*(Suppl 1), S41–S53.

Needleman, H. L., McFarland, C., Ness, R. B., Fienberg, S. E., & Tobin, M. J. (2002). Bone lead levels in adjudicated delinquents: A case control study. *Neurotoxicology and Teratology, 24*(6), 711–717.

National of Institute on Drug Abuse (NIDA) (March, 2007). *Drug abuse is preventable*. Accessed at http://www.drugabuse.gov/publications/topics-in-brief/drug-abuse-prevention on January 20, 2014.

New York State Commission on Drugs and the Courts. (2000). *Confronting the cycle of addiction and recidivism: A report to Chief Judge Judith S. Kaye*. Retrieved on December 16, 2013 from http://www.nycourts.gov/reports/addictionrecidivism.shtml

Piquero, A. & Piquero, N. (2002). Criminology and criminal justice research: Methods. *Encyclopedia of Crime and Justice*. Retrieved February 16, 2014 from Encyclopedia.com:http://www.encyclopedia.com/doc/1G2-3403000080.html

Potter, R. H., & Akers, T. A. (2013). Epidemiological criminology and violence prevention: Addressing the co-occurrence of criminal violence and poor health outcomes. In A. M. Viens, J. Coggon, & A. S. Kessel (Eds.), *Criminal law, philosophy and public health practice*. United Kingdom: Cambridge University Press.

Potter, R. H., & Rosky, J. W. (2013). The iron fist in the latex glove: The intersection of criminal justice and public health. *American Journal of Criminal Justice, 38*(2), 276–288.

Reingle, J. M., Jennings, W. G., & Komro, K. A. (2013). A case-control study of risk and protective factors for incarceration among urban youth. *Journal of Adolescent Health, 53*(4), 471–477.

Satcher, D. (1995). Violence as a public health issue. *Bulletin of the New York Academy of Medicine, 72*(1), 46–56.

Sullivan, L. (1991). Violence as a public health issue. *Journal of the American Medical Association, 265,* 2778.

Trickett, E. J., & Beehler, S. (2013). The ecology of multilevel interventions to reduce social inequalities in health. *American Behavioral Scientist, 57,* 1227–1246.

U.S. Bureau of Justice Statistics (2011, April). *Annual crime in the United States. Uniform Crime Report*. http://bjs.ojp.usdoj.gov/content/glance/tables/drugtab.cfm

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). *Prevention Science*, *2*, 165–192.

Whoriskey, P. (2012, November 24). As drug industry's influence over research grows, so does the potential for bias. The Washington Post. Retrieved from http://www.washingtonpost.com/business

Wilson, D. B., Mitchell, O., & MacKenzie, D. L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology, 2*(4), 459–487.

Part II
Quantitative Approaches

3 Randomized Controlled Trials in Substance Abuse Treatment Research: Fundamental Aspects and New Developments in Random Assignment Strategies, Comparison/Control Conditions, and Design Characteristics

James A. Swartz

3.1 Introduction

A *Los Angeles Times* article, published in November 2008 and entitled the "30-Day Myth", describes how the majority of residential substance abuse treatment programs—excluding self-help programs such as Alcoholics Anonymous—were designed to be completed in about 30 days (Roan 2008). The article indicates that the 30-day limit was arbitrary and largely based on insurance limits rather than on evidence supporting the effectiveness of this program length. It goes on to describe how, based on the consistent findings of recent scientific studies, more and more programs including one offered at the well-known Betty Ford Center, had lengthened treatment to 90 days. These studies were widely interpreted as establishing 90 days as the recommended, and scientifically based, minimum length of stay for achieving lasting drug treatment results. Treatment programs, such as those offered by The Betty Ford Center, followed suit and adjusted their offerings to provide "evidence-based" treatment consistent with these research findings. In fact, a report still available on the web site of the Better Ford Center cites National Institute on Drug Abuse (NIDA) funded researchers as indicating the "new gold standard" is 90 days of treatment (Betty Ford Center 2008).

And, in fact, the *Times* article is correct that over the three previous decades, studies had found that clinically significant treatment benefits such as abstinence from substance use and reduced relapse rates more likely occurred if a person spent at least 90 days in treatment (Fletcher et al. 1997). A succession of large-scale, federally funded studies, beginning with the Drug Abuse Reporting Program (DARP) in the early 1970s, the Treatment Outcome Prospective Study (TOPS) in the early 1980s, and the Drug Abuse Treatment Outcome Studies (DATOS) in the early 1990s were impressive in both their scope and consistency of findings associating longer lengths of treatment with better outcomes. Each of these longitudinal studies followed thousands of treatment participants enrolled in hundreds of federally funded treatment programs across varying modalities (e.g., methadone maintenance, therapeutic community, outpatient drug-free, etc.). As documented by the *Times* article and echoed by many other reports still available on the Internet as well as in the scientific literature, these findings have shaped both the processes and durations of drug

J.A. Swartz (✉)
Jane Addams College of Social Work, University of Illinois at Chicago, Chicago, IL, USA
e-mail: jaswartz@uic.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_3

treatment programs in the United States with the goal of increasing treatment retention in order to achieve better outcomes (Fletcher et al. 1997; Hubbard et al. 2003).

As influential as these studies have been, however, they had one very important methodological limitation; namely, they employed naturalistic, observational designs in which patients self-selected in and out of treatment.[1] In other words, participants largely determined their length of stay in treatment, up to the maximum duration provided by the program. This particular design characteristic was true of all three aforementioned national studies. In effect, each of these treatment programs replicated the basic quasi-experimental one-group pretest posttest design, illustrated in Campbell and Stanley's (1963) still authoritative explication of causal inference in experimental, quasi-experimental, and nonexperimental research designs as:

O-X-O

In this schematic, the first O, standing for observation, represents the measurement of participant characteristics at baseline, while the X represents the treatment intervention, and the second O represents the measurement of participant characteristics at some follow-up point. The replication of this design across many treatment programs and the consistent findings across programs that treatment length of stay is strongly associated with better outcomes and that at least 90 days in treatment are needed to produce clinically meaningful (or at least statistically significant) behavioral changes do lend some weight and credibility to the still frequently cited conclusions from DATOS-derived studies.

However, the nonexperimental design of the research (i.e., the lack of randomization and a priori comparison or control groups) means that despite the thousands of participants observed, the process and outcome measurements employed, and the follow-up rates obtained (>70%), there remains (or should remain) considerable doubt as to whether it was the 90-day treatment length per se that resulted in the observed clinical changes or whether it was simply that, for example, something about participants who opted to remain in treatment for that length of time.

It is exactly the confounding of participant characteristics with treatment retention and outcome in observational studies of substance abuse treatment—no matter how large in the scale—that leaves in question the extent to which treatment outcome(s) of interest are contingent on length of stay versus the characteristics of participants (and programs). Nor does it matter how well or how carefully baseline participant characteristics are measured and controlled for through a methodological (e.g., matching) and/or statistical (e.g., inclusion of covariates in the analytic models) technique. No other technique can provide an equivalent degree of confidence in the internal validity of a study compared with random assignment to treatment condition (Kao et al. 2008).

The *post hoc* comparison of people who complete or remain in drug treatment past a certain period of time with those who do not complete or drop out of treatment prematurely is subject to what is called an attrition effect (Shadish et al. 2002), one of a number of plausible threats to the internal validity of observational studies such as DARP, TOPS, and DATOS. The internal validity of a study refers to the degree to which the results can be attributed to the independent variable to the exclusion of all other possible explanations. Threats to internal validity are competing factors sometimes called "confounds" that, owing to the methodological shortcomings present in any study, could have caused the changes seen between the experimental groups. Studies high in internal validity, such as well-designed and conducted randomized

[1]The acronym RCT can also be used to refer to randomized clinical trials when the subject area of the study is to compare clinical interventions using a randomized design. Most of the studies referenced in this paper are, in fact, randomized clinical trials for this reason. We use the term RCT in a general if technically incorrect sense to refer to both randomized controlled and randomized clinical trials throughout the chapter. All of the examples given, however, are randomized clinical trials that compare the effectiveness of one or more treatments.

controlled trials (RCTs), offer the most protection against these alternative explanations of the observed study effects and, for this reason, are considered the strongest research design for establishing the *efficacy* and *effectiveness* of a clinical intervention (Blanco et al. 2013; Shadish et al. 2002).

In the case of the DARP, TOPS and DATOS findings under discussion, the independent variable was length of stay in a drug treatment program. However, because length of stay was not under the direct control of the experimenters in any of these studies and, consequently, because treatment participants were not randomly assigned to be in treatment for different lengths of time, it is possible the differences in outcomes attributed to varying lengths of stay are due as much to unmeasured differences among participants who remained in treatment longer as compared with those who chose to leave or were terminated from treatment prior to completing the full program of services. Even more tenuous, given the likely presence of attrition-related factors, are the conclusions on which the assumed gold standard of 90 days was based.

In fact, a study carried out in the United Kingdom around the same time as DATOS came to a very different conclusion using a large-scale, nonrandomized observational design but a different method of data analysis that sought to determine the optimum length of stay for inpatient and residential rehabilitation treatment according to treatment type and drug abuse characteristics (Gossop et al. 1999). The planned time in treatment, much like the US studies, varied by the specific treatment program and was not under the control of the research team. The UK researchers also reported a strong association between time in treatment and improved outcomes; for community-based residential treatment, they found a 90-day length of stay to be critical for achieving improved outcomes in community-based residential care. However, a critical threshold of 28 days for inpatient care was also noted. As longer time in treatment has been consistently linked with favorable outcomes in both their own and many other studies, the authors noted that

> *…many of the factors that predict treatment retention are the same as those that predict improved outcomes*. [emphasis added]… Time in treatment is a complex measure and one which should, in many respects, be regarded as a proxy indicator of other factors. The findings regarding treatment threshold may reflect the tendency of the more motivated patients to stay longer and engage better with treatment. Clients who actively participate in the programmes and make cognitive and behavioural changes during treatment achieve superior outcomes to others who stay for comparable periods but who do not make such changes (Gossop et al. 1999, p. 95).

The National Treatment Improvement Study (NTIES; Gerstein et al. 1997), yet another large-scale observational study conducted in the mid-1990s, also concluded that optimal lengths of stay varied by program modality. The NTIES likewise noted that improvement followed a linear course, with longer times corresponding to more improvement, but that there was no specific threshold below which no or minimal improvement was observed. Interestingly, studies based on NTIES data found that for some types of programs, longer durations of stay were actually counterproductive (Zhang et al. 2002).

It seems then that the 90-day "gold standard" for minimum length of stay in drug treatment rests on a methodologically tenuous foundation and may neither be golden nor standard for many treatment clients for whom shorter lengths of stay would produce lasting and clinically important changes, or for others for whom longer lengths of stay would be required to effect clinically meaningful and durable change. It is also likely, as the UK researchers found, that different treatment modalities require different minimal lengths of stay to achieve measurable and lasting clinical effects.

3.2 Randomized Control Trails and Substance Abuse Research

More confidence in the 90-day threshold would have been obtained if studies of the effects of length of stay had been conducted using an RCT design whereby participants were randomly assigned to control/comparison groups that

systematically varied by length of treatment stay, operationalized as an explicit independent variable. Only then can investigators begin to isolate participant (and program) factors related to both retention and outcome from the effects of treatment duration. This hypothetical RCT would also benefit from an intention-to-treat analytic framework (Pagoto et al. 2009), whereby participants are analyzed according to their initial control/comparison group assignment regardless of whether they completed the program. Such an approach, as opposed to comparing dropouts and completers *post hoc*, has been utilized in numerous evaluation studies of drug treatment outcomes (e.g., Fernández-Montalvo et al. 2008; McMahon et al. 1999), with all the attendant problems for causal inference to which that approach leads. The usefulness of an intention-to-treat analysis is dependent on achieving high sample retention rates for the study, which is another important issue when conducting randomized or quasi-experimental longitudinal studies (Del Boca and Darkes 2007b; Scott 2004).

A well-done RCT minimizes potential participant differences that are associated with both treatment retention as well as clinical outcomes across experimental groups. Moreover, and this is one of the main if not the main strength of RCT designs, *all relevant participant differences (confounders) need not be known to the researcher in order for their effects to be controlled across experimental conditions*. In theory, we can never know completely what participant characteristics, or interactions between participant and treatment program attributes, might affect retention rates and/or outcomes, even if previous research has identified important participant-related factors, such as cognitive deficits and antisocial personality disorder (see Brorson et al. 2013; Campbell et al. 2015). For this reason, random assignment of participants to a treatment condition is superior to other strategies, such as matching participants across conditions on known characteristics or careful measurement at baseline followed by statistical adjustment of possible confounders or covariates during analysis (Albert 2013; Blanco et al. 2013). Such strategies control only for known issues/confounders.

Perhaps the lack of RCTs within the context of the succession of the DARP, TOPS, and DATOS studies, despite the rich patient and practice information they provided, was one reason that studies of similar scope and design were not funded after the 1990s. Instead, the National Drug Abuse Treatment Clinical Trials Network (CTN), a multisite consortium of leading substance abuse researchers and community-based drug treatment providers, was created for the purpose of carrying out coordinated studies of promising new drug treatment protocols, improving the translation of research findings into practice, and advancing our understanding of the causes and clinical course of drug addiction (Retrosen et al. 2002). The structure of the CTN, in addition to the prospective and interventionist nature of the research (specific treatment protocols, eligibility criteria, and data collection methods, etc.), has fostered the use of multisite RCTs to study drug treatment over the past decade (Carroll et al. 2011). Studies conducted as part of the CTN as well as conceptual work on the methodological aspects of conducting RCTs generally (Mercer et al. 2007; Wolff 2000) and in drug treatment studies specifically (e.g., Campbell et al. 2012; Carroll and Rounsaville 2007; Hedden et al. 2006; Humphreys et al. 2013) have significantly advanced our knowledge of best practices with respect to carrying out RCTs.

Conducting an RCT is much more involved than determining how random assignment (or allocation) to a condition will be implemented as there are multiple methodological determinations to consider prior to and following random assignment. As enumerated by Machin and Fayers (2010), RCT design steps include (but are not limited to): (1) determining the research question, (2) participant selection, (3) choice of interventions, (4) choice of design, (5) allocating participants to the study conditions (randomization), (6) conducting the assessments (which includes when and by whom assessments are conducted), and (7) analysis and reporting.

In the remainder of this chapter, we consider a number of these steps that strongly influence research design and, hence, are especially relevant to a discussion of the design of RCTs in the context of substance abuse research; these include: (1) determining the research question, (2) participant selection, (3) choice of design, and (4) randomization strategies. We believe the interests of the investigative team, the particular population of substance abusers they will be studying, and the clinical settings of interest will largely determine the choice of interventions. Measurement issues in substance abuse research and statistical analyses are covered in other chapters in this volume (see Perron et al. 2017; see also Del Boca and Darkes 2007b; Tiffany et al. 2012).

A comprehensive discussion of a complex topic such as RCTs is not possible within the scope of a single book chapter. Instead, this chapter focuses on the basic design elements emphasizing standard practice, as well as recent RCT innovations and advances within the context of substance abuse treatment research, many of which are drawn from the extensive list of research done within the National Drug Abuse Treatment (CTN). Recently completed and ongoing CTN studies have employed one or more of the design characteristics under consideration, with some studies reflecting the cutting edge of substance abuse treatment and translational research.

Much of the material presented is relevant to quasi-experimental studies as well as RCTS, with the exception of randomization strategies. For instance, quasi-experimental designs still involve selecting who will participate, the nature of the control/comparison groups and the kind of treatments that will be compared. In lieu of randomization, a quasi-experimental design often uses some kind of matching strategy, either at the clinic or participant level, though participant selection and their subsequent allocation to condition is often less stringent. Quasi-experiments are done because they are the best design possible given a certain setting and/or resources. Valuable information is developed from quasi-experimental studies and they have an important place in substance abuse treatment research (Del Boca and Darkes 2007a, b); the relative merits and disadvantages of a wide variety of quasi-experimental designs that would be applicable to this content area have been well documented (Del Boca and Darkes 2007a, b; Shadish et al. 2002).

There are many resources available for the reader interested in learning more about the design and implementation of RCTs in the area of substance abuse treatment. A full list of CTN-related publications is available on the Clinical Trials Network Dissemination Library web site at: http://ctndisseminationlibrary.org/. For more detailed information on conducting RCTs, there are many good introductory, intermediate, and advanced texts, such as those already cited by Machin and Fayers (2010). Solomon et al. (2009) text on conducting RCTs to study community-based psychosocial interventions is a succinct introductory text that is especially relevant for RCTs of substance abuse treatment, which is most often community-based. Other good resources include Shadish et al. (2002) already referenced work, Friedman et al. consideration of the practical aspects of conducting both randomized and nonrandomized clinical trials (1998), and the National Institute of Health's (NIH) new and continually evolving web-based resource, *NIH Collaboratory's Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*.[2]

3.2.1 Determining the Research Question

> It's always best to start at the beginning.
> **—Glinda, the Good Witch of the North (1939)**

The very beginning of any study, including an RCT, involves the straightforward issue of determining the research question. Complications can arise in settling on a clearly defined research question, however, which is oftentimes affected by

[2]Available online at: http://sites.duke.edu/rethinkingclinicaltrials/.

financial resources, time constraints, staffing, training staff on new protocols, and other pragmatic issues that can affect the scope of the study. The importance of this decision is underscored by the fact that it has repercussions for all subsequent design and implementation steps.

We assume for the purposes of this chapter, that the research question in the area of substance abuse treatment research involves determining whether a new and relatively untested treatment or prevention protocol, either alone or in combination with one or more other treatment/prevention protocols (see Brigham et al. 2009), produces more positive outcomes in treatment clients than existing treatments (Hser et al. 2014). Close variations on this theme include: (1) modifying the delivery of an existing treatment or prevention protocol, such as when an in-person, clinic-based intervention is adapted for administration via computer, the Internet, or a portable device "app" (Bewick et al. 2010; Carroll et al. 2008), or (2) delivering an intervention originally developed and established for use with one particular population to a different population, such as testing the use of medications for treating opioid addiction for treating alcoholism (Pettinati et al. 2011). We also briefly consider arguments that this focus on comparative effectiveness research has led to a blind alley of sorts, and that the substance abuse treatment RCTs should shift to examine somewhat different research questions (Orford 2008).

The extent to which an intervention being studied has been previously investigated largely determines the *stage* of the clinical trial. The clinical trial stage will, in turn, determine the broad parameters of the study protocol. The stage of a study refers to the sequencing of a set of closely related studies that build upon each other and are designed to test an intervention in a variety of contexts that are close approximations of real-world conditions. Each study in the progressive sequence is called a stage, of which, according to the Stage Model of Behavioral Therapies Research, developed by the National Institute on Drug Abuse, there are three in substance abuse treatment clinical trials research (Carroll and Rounsaville 2007; Rounsaville et al. 2001).

3.2.2 Stage I Trials

Typical research questions addressed in stage I trials have to do with whether the intervention is feasible for a given population and context, in addition to how the intervention is best structured (e.g., number and duration of contacts, delivered face-to-face, through the Internet, by peers, medical personnel, etc.) to achieve the intended clinical goals. In the case of trials involving pharmaceuticals, issues related to safety, such as dosage level and administration routes, might also be considered. Stage 1 studies can involve designing the intervention or adapting an existing intervention for a new population, related development work, such as the creation of training and implementation manuals, as well as procedures for assessing the fidelity with which the intervention has been implemented, the determination of what measures should be used, and when to best assess outcomes.

Stage I trials are usually small-scale and enroll a limited number of carefully selected participants (Brigham et al. 2009). They may or may not involve random assignment to condition, as the effectiveness of the intervention to determine effect size relative to comparison conditions is not of central importance in these studies. For those familiar with NIH/NIDA grant awards, small-scale, time-limited studies funded under the R03, R21, or R34 funding mechanisms for developing and pilot testing are often stage I trials.

A recent example of a stage I trial is a study of low birth weight infants evidencing neonatal abstinence syndrome (NAS) as a result of their mothers using methadone during their pregnancies (Bogen 2011). This study sought to determine if early caloric enhancement could reduce the morphine doses and shorten hospital stays as compared with infants on standard formula. This previously untested intervention was then evaluated for adequacy of recruitment, protocol feasibility, and intervention efficacy.

3.2.3 Stage II Trials

If an intervention proves to be feasible, safe, and provides at least some indication of having beneficial effects relative to a control condition, it can be more rigorously evaluated in a stage II trial. Stage II trials test the *efficacy* of the intervention and, in doing so, often utilize random assignment to treatment condition designs. An intervention's efficacy is assessed using highly controlled conditions, as the researchers have almost total control in determining the clinical setting; the intervention to be tested is implemented using well-defined standards and is often manualized for this purpose. Clinicians involved in the trial are well trained in the treatment protocol, continuously or frequently monitored, and are retrained throughout the course of the study as needed to insure fidelity to the intended treatment model. Furthermore, participants are selected according to detailed and exclusive eligibility criteria. For instance, a stage II trial of an intervention designed specifically for opioid addicts might exclude poly-substance abusers, as well as anyone with a co-occurring mental illness, even though these conditions might be common in among individuals addicted to opioid substances.

At least one of the control conditions in a stage II trial typically involves *treatment as usual* (*TAU*), which should also be carefully monitored for fidelity throughout the trial (Baer et al. 2007). The TAU condition reflects the status quo, or the current standard for treating the condition that is the target of the new intervention, which is not always as straightforward as it might seem. In multisite trials, for example, what is "usual" at one site might be very different from what is "usual" at another, as might the effectiveness (Nunes et al. 2010). The corresponding research question in a stage II trial is whether the intervention shows improved (or at least equivalent) clinical outcomes relative to the TAU condition, under as close to ideal circumstances as is experimentally feasible to approximate. Effect size estimates in stage II trials can be used for power calculations for stage III trials. Unlike stage I trials, stage II trials do typically employ some form of random assignment to condition and, if possible, are implemented in a double-blind fashion, meaning that neither the clinician nor the participant (or the person administering the assessments) knows the assigned treatment condition.

A recent example of a stage II trial in substance abuse treatment research involved assessing the efficacy of injectable extended-release naltrexone for patients with higher severity alcohol dependence (Pettinati et al. 2011). Previous work suggested that naltrexone was only effective for patients with less severe alcohol dependence. The multisite study enrolled 624 alcohol-dependent participants, who were randomly assigned to either a placebo or experimental treatment condition that included a monthly injection of extended-release naltrexone. The results indicated that participants receiving the extended-release formulation of naltrexone had significant reductions in the number of heavy drinking days as compared to participants in the placebo condition.

3.2.4 Stage III Trials

Once the efficacy of an intervention has been demonstrated in ideal, tightly controlled conditions, its *effectiveness* is assessed in a stage III trial. For instance, the next step in the evaluation of extended-release naltrexone will likely be a Stage III trial, which employs less controlled conditions in order to determine if naltrexone continues to be an effective treatment for clients with more severe alcohol dependencies and, if so, the duration of the effects post trial. As this example implies, these usually large-scale trials test an efficacious intervention in "real-world" settings, staffed by clinicians seeing clients with substance use disorders who are less highly screened for study enrollment as those in stage II trials. For example, whereas participants in a stage II trial may be screened out if they have a co-occurring mental illness or abuse multiple drugs, they will be included in a stage III trial because these kinds of clients are common in community-based programs. The reason for

expanding the inclusion criteria is that if the new treatment is only effective in a narrowly defined group of treatment clients, its chances for adoption in clinical settings are greatly diminished. Community-based programs see many different kinds of clients whose pattern and severity of substances misused vary considerably, as do the mix and severity of ancillary, health and psychosocial conditions with which they present. Adopting a new type of treatment is expensive and time consuming to implement, owing to training and staffing costs. To justify these investment costs, any new treatment must have been demonstrated effective in less rigorously controlled and monitored conditions with a relatively diverse client case-mix, hence the rational for Stage III effectiveness trials.

The research question addressed by stage III trials then is how effective is an intervention with demonstrated efficacy, in the context of treatment delivered, under conditions that more closely approximate real-world circumstances with more typical treatment clients. A related question of increasing importance and a focus of the National Drug Abuse Treatment CTN is how treatments established as efficacious/effective can best be translated/adopted into practice (Tai et al. 2011). It has been known for some time that despite the evidence-supported therapies (ESTs), substance abuse treatment programs have been slow to adopt them, or the ESTs have not been well implemented (Carroll and Rounsaville 2003; McGovern et al. 2004). Stage III trials can include an examination of this issue in what is called "translational research", which seeks to develop strategies for improving the adoption and implementation of ESTs (Woolf 2008).

A study of the effects of giving prize-based incentives to stimulant abusers in outpatient treatment illustrates a stage III clinical trial. Petri et al. (2005) recruited 415 cocaine or methamphetamine abusers entering outpatient treatment across 8 community-based clinics. Within each clinic, participants were either randomized to TAU or to TAU plus an abstinence-based incentive condition that involved drawings for chances to win prizes for submitting drug-free urine samples. Because this was a stage III RCT, TAU was allowed to vary across clinics. Study eligibility criteria were minimal (e.g., recent cocaine or amphetamine use by self-report or urinalysis) to enroll a broadly representative sample. Drawings for prizes contingent on submitting negative urines were used rather than vouchers; although the effectiveness of using a voucher to pay for each drug-free urine test has been documented in improving treatment retention and outcomes, vouchering is expensive as compared to drawing-based incentives. The expense of vouchering has been a barrier to adopting incentive-based interventions. By using a less expensive form of contingency management, the study sought to not only assess the effectiveness of incentive-based drawings in community clinics, but also to address one of the main issues preventing the wider adoption of contingency management procedures: cost. The study found that participants in the incentive condition, relative to those in the TAU condition, were retained in treatment longer, attended more treatment sessions, and had a higher number of consecutive treatment visits with confirmed abstinence.

Strictly speaking, this study of contingency management for stimulant abusers represents what Carroll and Rounsaville (2003) have described as being a **hybrid efficacy–effectiveness study**. The hybrid design, as the name suggests, combines aspects of stage II and stage III clinical trials and has become the norm across CTN studies. It retains elements of the stage II designs, in that randomization to conditions, close monitoring of treatment fidelity, and a well-defined treatment protocol are utilized. Stage III design elements include a broad inclusion criteria for participant selection, conducting the trial across multiple treatment settings, various terms of treatment, and other program characteristics.

The relevance of this kind of hybrid design for translational research suggests it, or close variants, will be increasingly used to evaluate new and existing treatments for substance abuse, both within and without the CTN context, given the growing emphasis on translational research and evidence-based interventions in clinical practice

settings. Hien et al. (2009) study, comparing the effects on substance use for women with co-occurring PTSD and substance use disorders, randomly assigned participants in either a Seeking Safety intervention or a women's health education comparison condition, which is yet another good example of a hybrid efficacy–effectiveness design.

3.3 Should the Research Focus on Treatment as a Technique in Substance Abuse Treatment RTCs Be Reconsidered?

As the above discussion implies, most substance abuse treatment RCTs in the CTN have compared the effectiveness of new interventions to TAU. As such, the primary research question in these studies is whether some new treatment alone, or the respective treatment's combination with TAU, proves to be more effective than TAU alone, or TAU in combination with yet another treatment. When multiple interventions are compared against each other and/or TAU, the study is called **comparative effectiveness research** (CER). In a less frequently used variation of a CER study, a **non-inferiority trial** design is used to determine if a new treatment is equivalent (i.e., equally effective) to an existing treatment (Wallace et al. 2013). A non-inferiority trial is especially useful when evaluating a new intervention that is briefer or less intensive than TAU, or delivered in a novel fashion, such as via computer.

In spite of their methodological rigor, RCT CER studies of substance abuse treatment have often yielded disappointing findings. Most studies have found, at best, moderate effect sizes for new treatments (Carroll et al. 2011; Nunes et al. 2010; Orford 2008) in comparison to the control conditions. Carroll et al. (2011), in their summary of the first 10 years of the National Drug Abuse Treatment CTN, noted that the lack of studies showing large effect sizes could be because RCTs in the CTN have generally used active control conditions, such as TAU or TAU plus another intervention, as opposed to inactive conditions such as wait list controls. It could be argued that, due to unexpected effectiveness of existing treatments, large effect sizes for new treatments are harder to demonstrate. Furthermore, the intense monitoring of treatment, including the TAU condition, could itself impact effectiveness by way of the Hawthorne effect, thereby boosting the effects of TAU (Carroll et al. 2011).

In commentary on the current state of CER substance abuse treatment research, Orford (2008) offers a number of other reasons as to how different interventions have yielded similar outcomes; one is simply that the interventions might not really be all that distinct when compared directly (the "equivalence paradox"). He believes that treatment studies have focused too much on technique—what Del Boca and Darkes (2007a) term the "technology model" of treatment—and not enough on the therapeutic relationship:

> Some of those who have written about psychotherapy more generally have referred to this as the 'drug metaphor', implying that treatment is seen, like a medication, as a piece of technology that requires only therapist skill and efficiency and patient compliance in order to be delivered effectively. That is a powerful model, but it may be inappropriate. There have always been voices raised against it, suggesting that the essence of psychological treatment is not a technique but rather the therapist–client relationship (Orford 2008, p. 2).

The findings from a recent RCT within the National Drug Abuse Treatment CTN support this observation. Campbell et al. (2015) randomly assigned 234 stimulant abusers, across 10 outpatient treatment programs, to both group and individual 12-Step Facilitation (TSF) plus TAU, or to TAU alone. Primary independent variables were the therapeutic alliance and therapist competence, as well as treatment fidelity and adherence. The study found that therapeutic alliance, therapist competence, and treatment fidelity were associated with longer treatment retention and better outcomes, yet adherence was not. They concluded that future studies should focus on the therapeutic alliance and on general therapist skills, a conclusion with which Orford would most certainly agree.

Some of Orford's other suggestions for shifting the focus of substance abuse treatment CER studies include:

1. Giving stronger consideration and emphasis to unaided change, whereby individuals with addictions improve outside of any professional or therapeutic context. Studying both unaided as well as clinically driven change, Orford argues, would provide better insight into the change process generally.
2. Studying drug addiction as a chronic condition by not restricting the focus of studies to short-term (e.g., 12-month) outcomes. Carroll et al. (2011) have also suggested that longer term trials are needed, despite the expense, to develop adaptive interventions and better understand addiction over the life course (see Hser et al. (2015) for an example of a longer term outcomes study of opioid-using addicts in medication-assisted treatment).
3. Understand and incorporate the therapists "tacit theories" of change, as well as the patient's view of treatment and the change process. Orford takes this a step further by recommending greater utilization of qualitative methods and mixed methods-based studies to gain detailed knowledge of participant views, as opposed to a heavy reliance on quantitative methods influenced by a medical bio-behavioral model of treatment research. The National CTN has taken steps in this direction by making community treatment providers full partners with researchers in order to determine the design and conduct of research studies so that "practice-relevant" questions, in addition to clinicians' perspectives, are incorporated into treatment elements (Tai et al. 2011).

3.3.1 Participant Selection

Many beginning graduate students make the fundamental mistake of confusing random assignment with random selection. They are also inclined to include both random selection and assignment in their respective study designs; although they share a common language of probability in their defining characteristics, the purposes of these two fundamental research techniques are distinct. It takes some reflection to appreciate the difference between having an equal (or known) chance of being selected for a study and having an equal (or known) chance of being assigned to one of the experimental conditions. Random selection has to do with obtaining a representative sample, and hence is concerned with increasing the external validity, or generalizability of the study. Random assignment—sometimes termed random allocation—to condition has to do with increasing the internal validity of the study and the associated degree of confidence in the conclusions drawn about the study hypothesis and treatment effects.

Despite the benefits of enhancing sample representativeness, random selection is not often used in RCTs; there are a number of reasons for this. First, RCTs, especially stage I and II trials, are more concerned with maximizing the internal validity of a study as opposed to its level of generalizability. Even in stage III effectiveness trials, the generalizability of the intervention is not established by random selection, but rather by broadening participant inclusion criteria, conducting the study in multiple, community-based drug treatment settings, and by having the intervention delivered by multiple clinicians.

Second, even if it were feasible to implement random selection as part of determining who is enrolled in a study, not much is gained by doing so. If the sampling frame from which the sample is selected consists of all admissions to a specific treatment program, this becomes the population to which the study findings are generalizable. As clinical trials research is concerned with making broader inferences about the effectiveness of a particular treatment for clients with the condition being treated, such as the case with drug addiction, the importance of generalizing back to the treatment population within a specific program is largely irrelevant. The real question, however, is how well do interventions work for those with the condition in question if they receive the

intervention within any similarly structured program, implementing the same set of clinical techniques.

Third, obtaining participants for clinical trials, whether randomized or not, is time-consuming. Accruing participants into a study to fulfill target numbers, driven by statistical power considerations, is determined by clinical capacity and takes time, particularly when there are a large number of eligibility criteria applied to participant selection. The candidate pool can become small depending on the size of the program and number of admissions over the study time period. Adding random selection, in addition to eligibility criteria, could considerably lengthen the recruitment period and place additional strain on study resources, thereby limiting the follow-up period.

Finally, unlike survey studies where a sampling frame is drawn up and a sample selected through some probabilistic method (e.g., simple random sample, stratified random sampling, etc.), sampling frames often lack feasibility for clinical trials as most recruit people who are newly entering treatment, a practice that provides for equivalence with respect to the treatment exposure. It would greatly complicate matters and significantly reduce the internal validity of the study to include participants with varying amounts of time in and exposure to treatment at their point of entry into the study. For this reason, most RCTs in substance abuse research include eligibility criteria for participants who are new admissions to the treatment program.

Thus, with respect to RCTs, client eligibility characteristics and not random selection are the most important determinants of who is included in the clinical trial along with who agrees to participate in the study, the latter of which is generally not under the control of the experimenter. The specificity of the eligibility characteristics, in turn, is strongly associated with whether the study is a Stage II efficacy trial or a Stage III effectiveness/translational trial. Stage II trials have historically been very selective of participants; this is both a strength and weakness of these studies. On the one hand, having a clear set of participant eligibility criteria, along with a well-defined and carefully monitored intervention, enhances the study's internal validity. On the other hand, if the sample is so highly selective that it bears little resemblance to the population of individuals who seek treatment in non-research clinical settings, the studied intervention could be much less effective than demonstrated in the RCT or even completely inapplicable and ineffective (Humphreys et al. 2013; Kunz et al. 2008; Susikida et al. 2016).

A review of clinical trials evaluating treatment for drug dependence, for example, found that 26% of screened patients were ineligible and that another 32% of eligible patients refused to participate (Melberg and Humphreys 2010). A review of alcohol dependence treatment studies found that the eligibility criteria used would exclude upwards of 79% of patients seeking treatment (Blanco et al. 2008), as those enrolled in clinical trials, relative to all treatment seekers, were more likely to be dependent on one specific substance and not poly-substance abusers, had less severe dependencies, were less likely to have co-occurring medical or psychiatric conditions, had higher motivation for treatment, and were less likely to be African American (Blanco et al. 2008; Carroll et al. 2011; Humphreys et al. 2005).

The issue of restricted participant selection and its impact on the generalizability and application of RCT research has been recognized and is being addressed by the National Drug Treatment Research CTN (Carroll et al. 2011). The majority of CTN studies conducted to date qualify as either Stage III clinical trials or hybrid efficacy–effectiveness trials, as described above. As such, they have used broader inclusion criteria during participant selection; more work needs to be done in this regard, however. For instance, participants with co-occurring psychiatric disorders have not been routinely included or specifically studied in all but a few CTN studies, despite the high prevalence of such conditions in the substance abusing population, particularly those seeking

treatment (Flynn and Brown 2008). As noted by Carroll et al. (2011), only two treatment studies for clients with co-occurring psychiatric disorders have been conducted by the National CTN. They note that this is due to the heterogeneity of this population "…in terms of issues such as variability in different Axis I and II disorders, drugs of abuse, illness severity, and use of psychotropic medications." Still, despite the methodological complexity it causes, it is important to not only study interventions specifically tailored for persons with co-occurring psychiatric and substance use disorders, but to also make co-occurring disorders an exclusionary criteria in RCTs of community-based substance abuse treatment. The high prevalence of mental illnesses among substance abusers means the continued exclusion of persons with co-occurring mental illnesses and substance use disorders from treatment effectiveness studies all but insures study findings will have limited applicability in most community-based settings.

3.3.2 Choice of Design and Control/Comparison Groups

The basic design for an RCT, commonly utilized in both substance abuse treatment stage III effectiveness research and stage II efficacy research, as mentioned above, is deceptive in its diagrammatic simplicity. Using the conventions to represent design alternatives in Shadish et al. (2002), the workhorse design for substance abuse treatment RCTs can be represented as[3]:

$$\begin{array}{cccc} R & O & X_a & O \\ R & O & X_b & O \end{array}$$

This design, termed the "alternative-treatments design with pretest," has two groups, each of which is randomly assigned prior to a treatment condition before baseline measurement. What qualifies this design as an RCT is, of course, the randomization to condition, as well as the inclusion of an explicit, concurrent comparison/control condition. It is these two critical elements that distinguish this design from the ones used in the DARP, TOPS, and DATOS studies, giving it greater methodological rigor and potential for higher internal validity.

In multisite trials, this design is replicated at each site, with site and therapists nested within site becoming additional factors to assess analytically, either as fixed or random effects, along with the effects of the intervention and other covariates of interest (see Feaster et al. 2011 for a discussion of fixed versus random effects). For instance, Ball et al. (2007) used an RCT design to compare motivational enhancement therapy and counseling for increasing treatment retention and reducing substance use. Participants were recruited from among those receiving care in one of five outpatient substance abuse treatment programs. The study found main as well as interaction effects for both program site and treatment conditions, with effect sizes varying by site and "inconsistent in direction." The authors noted that site effects such as those found in their study are common in multisite studies.

Baseline measurements are not required but are highly desirable. *Post hoc*, they help the investigator determine if the randomization has successfully balanced the experimental groups on key elements related to outcomes. If the groups are unbalanced in terms of important covariates, the baseline data can be used to adjust statistically for these differences during analysis. Another general benefit of including a baseline measurement, regardless of whether randomization is used, is that it provides information for determining the amount of change at outcome relative to baseline. If baseline measurement is done prior to random assignment, as in the following diagram, there is an additional benefit of being able to use the information to fine-tune the randomization with a procedure called **urn randomization** (see below).

[3]Arguably, this same design but without random assignment to condition is also the kind of design used most often in quasi-experimental research studies. In the quasi-experimental version, in place of random assignment, some other strategy such as matching is used to maximize the comparability of participant characteristics across the experimental conditions.

$$\begin{array}{cccc} O & R & X_a & O \\ O & R & X_b & O \end{array}$$

Another important advantage of randomizing after baseline assessment is that it is easier to blind the baseline assessors to the randomly assigned treatment condition of the participant being assessed, thereby minimizing any potential bias when the researcher conducting the baseline assessment is non-blinded. Regardless of when randomization occurs relative to the baseline assessment, however, potential measurement bias can be minimized if the assessors are blinded to the experimental condition of the participant being assessed.

Following random assignment, each group receives an intervention and then a second measurement is conducted at some point posttreatment assignment to compare outcomes and to determine the efficacy/effectiveness of the interventions under study. Although not typically diagrammed in models representing experimental designs such as this, measurement of treatment outcomes is often completed during treatment, as well as after formal treatment is completed; it is recommended that such "in-trial" assessments are routinely included. As treatment, as well as experimental attrition, inevitably occur over the course of a study, including multiple measurement points along the way reduces the potential for biasing the effects of missing data, allows for determination of the missing data mechanism (i.e., missing completely at random, etc.) and, using statistical techniques such as hierarchical liner modeling, participants with missing data can still be included in the analyses (Gibbons et al. 2010; Hedden et al. 2009; McPherson et al. 2015).

Not shown in the diagram is the intensive amount of assessment work that must be done by the investigative team during the intervention. In well-designed and conducted RCTs (or in any study of treatment interventions, whether randomized or not), a considerable amount of attention is given to training staff and measuring treatment fidelity (Baer et al. 2007). That is, there is very intensive evaluation of how the intervention is being delivered to insure it is being done as prescribed (i.e., that there is high treatment integrity). The necessity of delivering treatment as prescribed is one reason manualized treatments, where exact specifications can be enumerated and explained for the training and treatment teams, have come to predominate in RCTs. It is worth reemphasizing that it is important to remember, to assess the treatment fidelity of the TAU control conditions to insure that bias is not introduced by giving greater scrutiny to one condition over another.

There is no methodological disadvantage, and much to be gained by including multiple assessments post-intervention, such that the basic design model is extended longitudinally. Multiple baseline measurements can also be added, but designs employing these are rare in RCT studies (Shadish et al. 2002). The main disadvantages of multiple assessments are the time and financial resources needed to conduct an extended longitudinal study. An obvious advantage of longitudinally extended designs is that they allow for assessment of change over time through the use of growth curve modeling and related analytic techniques, such as growth mixture modeling. These designs also allow the investigator to assess the durability of treatment effects posttreatment discharge. For instance, a multisite RCT investigating the effectiveness of varenicline, in conjunction with cognitive behavioral therapy (CBT), for smoking relapse prevention among patients with schizophrenia or bipolar disease used an extended, longitudinal design with multiple weekly post-intervention assessments (Evins et al. 2014). Following an open-label trial on varenicline, participants who met abstinence criteria were randomly assigned to a double-blinded trial of CBT plus either varenicline or placebo. Participants were reassessed from weeks 12 through 76 of the study. The study found varenicline plus CBT to be superior to CBT alone after one year of treatment, with the beneficial effects maintained at 6 months following treatment conclusion.

Depending on the research question and trial stage, the investigator has wide latitude in

determining what treatment, or treatments, will compose the interventions used in the comparison conditions. Resources allowing, the basic design can be, and often is, extended to include more than one comparison group. Most often, however, at least one of the comparison group interventions will be TAU. The intervention under study can be administered as a stand-alone and completely self-contained alternative to TAU (Witkiewitz et al. 2014), as an adjunct to be delivered in addition to TAU (Marsch et al. 2014), or sometimes both.

It has become less common in substance abuse treatment research to use a control condition that receives no treatment at all. For one, the ethics of randomly assigning someone who is seeking treatment to receive no treatment, even if treatment is delayed, as is common in a waiting list control group or in an "attention control" condition where services less than TAU are provided, militates against the feasibility of doing so when TAU is available. A study that proposes withholding available treatment can and should meet with considerable resistance from an IRB. Second, and in stage III trials in particular, in order to justify the time and costs of adopting a new intervention that intervention should demonstrate greater or at least equivalent effectiveness, to the treatments already in use. The comparison of new interventions against so-called active treatment control conditions, such as TAU, is a conservative strategy, making it harder to find large treatment effects for the new interventions or combinations of interventions (Karlsson and Bergmark 2014). However, the importance of improving upon treatment that already exists would seem to dictate the use of designs where TAU is at least one of the comparison conditions.

An exception to having TAU as a comparison condition occurs in studies comparing the efficacy/effectiveness of two (or more) new interventions, either alone or in conjunction with TAU, as would be compared in a study comparing the effectiveness of buprenorphine versus clonidine for opioid detoxification (Ling et al. 2005), where neither condition could be considered TAU per se. A number of other variations, depending on how treatments are combined and compared across conditions are possible, each suited for addressing a specific research question and with differing strengths and weaknesses. Nunes et al. (2010) provide a full enumeration of possible designs by discussing the relative merits of each design based on the CTN RCT studies that had been conducted up to the time of publication.

3.3.3 Adaptive Designs

Over the past decade, **adaptive designs** have become more popular in clinical intervention research but remain uncommon in substance abuse treatment studies and in other applicable research areas limited by the complexity of implementation, as well as data analysis and cost. Adaptive designs use information collected from study participants at an early stage in the study to determine how an intervention should be modified, or a new intervention tried at a later stage in the study (Murphy et al. 2007). These modifications can include more intensive or additional care, if needed, by way of stepped care models or through re-randomization to a different treatment condition in what Murphy et al. (2007) termed a sequential multiple assignment randomized trial (SMART). Adaptive designs have the potential to better approximate how therapy is administered in real-world conditions where treatment clients can be switched from one type of treatment to another if the first treatment proves ineffective or there is a problem with adherence. Additionally, as noted by Coffey et al. (2012), this approach provides investigators with an attractive solution to address some initial design uncertainties over which treatment is most effective for the type of client that exists. Adaptive designs also make sense when considering the chronic nature of substance abuse, as well as the heterogeneity of the substance abusing population, moving the field away from a one-size-fits-all treatment model that predominated the early years of substance abuse treatment and remains common through the present day (Swartz 2012).

A substance abuse treatment study employing a SMART design was used to assess the effects of reinforcement-based therapy for pregnant women with opiate and/or cocaine substance use disorders (Jones et al. 2011). Participants were first randomized to traditional or reduced RBT and then assessed for compliance after two weeks via drug testing. Women in the two different conditions were then re-randomized to reduced RBT, traditional RBT, enhanced RBT, or abbreviated RBT, contingent upon their initial treatment assignment and compliance. The interested reader may be referred to the Methodology Center website at Penn State, which maintains a catalog of current and completed studies employing SMART designs across a variety of research areas, including substance abuse treatment.[4]

3.3.4 Randomization Strategies

At last we consider the heart of RCTs, strategies for randomizing participants to experimental conditions. The procedures described in this section are concerned with balancing participant characteristics across intervention groups. For a discussion of randomization procedures that balance numbers across treatment conditions, see Hedden et al. (2006).

The most basic form of random assignment, called **complete randomization**, utilizes a random numbers table or, more commonly now, a computerized algorithm to assign each participant to one of the experimental conditions without restrictions; random assignment is achieved by each participant having an equal and known chance of being assigned to each condition. In the simplest case, for which there is one experimental and one control condition, each new participant has a 50% chance of being assigned to either condition. If there are three experimental conditions, the odds of assignment to any one are 33%, and so on.

In many circumstances, a more elaborate strategy is unnecessary to insure participants across conditions are matched on both known and unknown factors that might influence outcomes other than the intervention(s) under study. In fact, for larger trials (e.g., $N > 400$), complete randomization works well in achieving balance on important participant characteristics between groups (Hedden et al. 2006). It is important how the random assignment strategy is implemented, and best practices suggest that randomization is either done by research staff that are not involved in assessing participants, or by clinical staff who will be delivering the treatment, both of which should be blinded to the assignment process and outcome (Machin and Fayers 2010). While it is difficult for treatment staff to remain blinded as to the treatment assignment in studies of non-pharmacological behavioral interventions, where possible, staff conducting participant assessments should always be blinded to the original treatment assignment as they conduct their assessments over the course of the study.

However, as noted above, random assignment is not foolproof and it is possible that even after random assignment, the study groups are imbalanced on important known attributes or even in the numbers of participants allocated to each condition, resulting in a threat to the study's internal validity known as *selection bias*. The likelihood of imbalance is greater when the number of participants in each condition is small and the population from which they are drawn is heterogeneous (Blair 2004). For this reason, strategies designed to better insure balance have been adopted for use in RCTs. One such strategy, *stratification with simple random assignment*, divides participants into groups (strata or blocks) based upon characteristics that could potentially influence the outcome being studied. Random assignment to experimental conditions is then applied to each stratum independently. This process insures that the experimental conditions will be balanced with respect to the variables that define the strata.

This strategy is most effective, however, when there are limited stratifying variables (Hedden et al. 2006; Machin and Fayers 2010). As the number of stratifying variables increases, so too does the possibility that there will be subgroups

[4] http://methodology.psu.edu/ra/smart/projects.

with very few cases available for randomization. For these cases, assignment to condition could become determinant (i.e., nonrandom). Hence, this strategy is only recommended when there is a small number of stratifying variables. For this reason, and for situations where balance among the experimental groups on more than a few variables is desired while preserving random assignment, dynamic allocation procedures, such as *covariate adaptive and urn randomization*, are used (Stout et al. 1994).

Dynamic allocation procedures, sometimes referred to as "minimization procedures" because they attempt to minimize the differences on known confounders between participants in the experimental and control conditions (Blair 2004), are akin to adaptive designs in that they use already collected information to make prospective decisions. They are also termed "restrictive" in the sense that they place balancing constraints on the randomization process (Hedden et al. 2006). Dynamic randomization procedures track the proportional composition of the experimental groups in terms of the variables on which balance is desired. The attributes of the next case to be randomly assigned are weighed against the existing proportions of already assigned cases, with a higher probability of assignment to the group that has the lowest proportion of cases with similar attributes. Thus, the probabilities of assignment to condition are continuously adjusted by the dynamic allocation procedure, usually implemented through a computer program.

Albeit more complicated to administer than non-dynamic random assignment procedures, the probability of imbalance (and hence selection bias) is lower with dynamic procedures (Stout et al. 1994), especially when sample sizes are small. The two methods converge when sample sizes are large, tilting the selection of a randomization strategy toward complete random assignment because it is simpler to implement and requires fewer adjustments during data analysis to compensate for the shifting probabilities of assignment to condition (Heden et al. 2006).

Given that covariate adaptive and other dynamic allocation procedures are probabilistic, it is still possible that the experimental groups will not be balanced on all covariates. In fact, as the allocation process unfolds, it is possible for the groups to become so imbalanced that the assignment of the next available case with a specific set of attributes becomes "deterministic" (i.e., predictable with high certainty). For this reason, studies that use urn randomization, or some similar minimization procedure, also include an additional probabilistic component to reduce the determinism of the assignment.

When selecting what variables to include in a dynamic randomization procedure, it is important to select those variables that are most strongly associated with the outcomes being investigated and to avoid including weakly or unassociated variables. For example, in an RCT studying the effectiveness of a brief alcohol intervention to reduce heavy drinking during smoking cessation treatment, urn randomization can be used to assign participants to treatment conditions: level of nicotine dependence, number of drinks consumed per week, and intention to change drinking while quitting smoking (Kahler et al. 2009). The selection of these specific variables for stratification was based on the previous smoking cessation studies showing each to be related to treatment outcomes.

3.4 Conclusions

> Whoever is seeking documentation of clinical practice needs to be critical enough to avoid the lure of the gold standard in assessing evidence, so as to not end up… with fool's gold.
>
> —A. Rosner (2002)

Hopefully this overview of RCTs as they pertain to substance abuse treatment research has provided a useful introduction to a complex topic. The methodology of RCTs continues to evolve rapidly with advances at every level, from selecting participants, determining the experimental conditions of focus, how participants should be allocated to conditions, to how the intervention fidelity will be assessed. Driving these changes is the need to develop more

effective interventions, including strategies for improving treatment (and study) retention, as well as those meant to insure that effective interventions are adopted and faithfully implemented in practice settings outside of the experimental context.

This chapter began with a consideration of RCT benefits relative to observational studies. Much of the ensuing discussion highlighted and described how implementation of an RCT achieves these benefits. It is fitting then, perhaps, to conclude with a consideration of the limitations of RCTs and point out that an RCT is not always a superior research protocol given a particular substantive area, research question, and clinical context. Moreover, a poorly conducted RCT could be far less valuable than a well-conducted observational, quasi-experimental study. As Grossman and MacKenzie (2005) have pointed out:

> To argue from the fact that RCTs have certain advantages, other things being equal, to the claim that the RCT is a gold standard, is like arguing that since being tall makes for a good high-jumper, it follows that a 6′ elderly drunkard with a spinal injury is bound to be a better high-jumper than a 5′ 11″ Olympic athlete. All things are never equal, and one has to consider many factors other than, in this example, the person's height. Just as being tall is often a good property for a jumper to have, the property of being an RCT is often a good property for a study to have, but it does not follow that anything that is an RCT is better than anything that isn't. (Grossman and MacKenzie 2005, p. 520)

RCTs do have limitations and are not without their critics. They are expensive and time consuming. As put succinctly by Shadish et al. (2002), "randomized experiments may not be desirable when quick answers are needed." The staffing requirements, as well as the technical expertise needed to successfully conduct a well-run RCT, have limited the pool of eligible investigators to those with considerable experience, as well as the ability to obtain large federally funded research grants. Often, multisite trials must be done to accrue enough participants and to study the intervention(s) in a variety of treatment settings. It can be difficult for new investigators with innovative ideas to obtain the funding needed to conduct RCTs. And hence, the cost, size, and complexity of RCTs may reduce innovation in an area of research that needs it considerably. For some of these reasons, alternative but still valuable study designs, such as the multiple baseline and case control studies, have been proposed as viable, less costly alternatives to RCTs that provide valid and valuable information on treatment effectiveness (Grossman and MacKenzie 2005; Hawkins et al. 2007).

In some settings, RCTs may simply not be possible. This is especially true, for instance, in studies of substance abuse treatment in criminal justice settings, such as in evaluations of drug courts or prison-based treatment programs. Judges, attorneys, and probation officers are unlikely to agree to allow a drug-dependent offender to be randomly assigned to a treatment condition. In such circumstances, a study that employs a case control or other matching strategy is the optimum available design choice.

Critics have also noted that RCTs might, by their very nature, produce findings that are inapplicable to real-world settings. For instance, Kaptchuk (2001) has commented that there is a possibility we need to consider a "Heisenberg Principle of [Human Experimental] Sciences whereby the act of setting up controls can alter the phenomenon sufficiently to yield quite different results." For example, in an RCT, does the level of scrutiny needed to assess treatment fidelity and the act of conducting multiple assessments with participants alter the very nature of the treatment being studied? Wolff (2000) has argued that RCTs might not be suitable for evaluating "socially complex services," such as substance abuse treatment, because in most trials the effects of the social environment, such as the communities in which trial participants reside and/or return to following treatment, are not taken into account. She also notes the selection biases and the non-representativeness of who is included in RCTs, as was discussed in the section on participant selection above. To an extent, as we have also noted, these kinds of selection biases have been reduced in the RCTs conducted in the National CTN.

Notwithstanding these legitimate limitations and criticisms, however, RCTs will continue to evolve and, at least for the foreseeable future, continue to be seen, when they are applicable to a substantive area and properly conducted, as providing a high evidentiary standard for evaluating treatment effectiveness. The challenge is to continue to refine the methodology to address the identified limitations while maintaining the essential elements of an RCT, delineated in this paper and in many of the provided references, which enhance the internal and external validity of these studies.

References

Albert, R. K. (2013). "Lies, Damned Lies…" and observational studies in comparative effectiveness research. *American Journal of Respiratory and Critical Care Medicine, 187,* 1173–1177.

Baer, J. S., Ball, S. A., Campbell, B. K., Miele, G. M., Schoener, E. P., & Tracy, K. (2007). Training and fidelity monitoring of behavioral interventions in multi-site addictions research. *Drug and Alcohol Dependence, 16,* 107–118.

Ball, S. A., Martino, S., Nich, C., Frankforter, T. L., Van Horn, D., Crits-Christoph, P., et al. (2007). Site matters: Multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology, 75,* 556–567.

Betty Ford Center. (March 1, 2008). *90 day treatment stay the new "gold standard"*. Retrieved from http://www.bettyfordcenter.org/recovery/programs/90-day-treatment-stay-the-new-gold-standard.php

Bewick, B. M., West, R., Gill, J., O'May, F., Mulhern, B., Barkham, M., & Hill, A. J. (2010). Providing web-based feedback and social norms information to reduce student alcohol intake: A multisite investigation. *Journal of Medical Internet Research, 12.* Advance online publication. doi:10.2196/jmir.1461

Blair, E. (2004). Gold is not always good enough: The shortcomings of randomization when evaluating interventions in small heterogeneous samples. *Journal of Clinical Epidemiology, 57,* 1219–1222.

Blanco, C., Olfson, M., Okuda, M., Nunes, E. V., Liu, S.-M., & Hasin, D. S. (2008). Generalizability of clinical trials for alcohol dependence to community samples. *Drug and Alcohol Dependence, 98,* 123–128.

Blanco, C., Rafful, C., & Olfson, M. (2013). The use of clinical trials in comparative effectiveness research on mental health. *Journal of Clinical Epidemiology, 66,* S29–S36.

Bogen, D. L. (2011). Randomized clinical trial of high vs. standard-calorie formula for methadone-exposed infants [NIDA grant proposal]. Retrieved from http://projectreporter.nih.gov/project_info_description.cfm?aid=8049612&icde=22651899

Brigham, G. S., Feaster, D. J., Wakim, P. G., & Dempsey, C. L. (2009). Choosing a control group in effectiveness trials of behavioral drug abuse treatment. *Journal of Substance Abuse Treatment, 37,* 388–397.

Brorson, H. H., Arnevik, E. A., Rand-Hendriksen, K., & Duckert, F. (2013). Drop-out from addiction treatment: A systematic review of risk factors. *Clinical Psychology Review, 33,* 1010–1024.

Campbell, B. K., Guydish, J., Le, T., Wells, E. A., & McCarty, D. (2015). The relationship of therapeutic alliance and treatment delivery fidelity with treatment retention in a multisite trial of twelve-step facilities. *Psychology of Addictive Behaviors.* Advance online publication. doi:10.1037/adb0000008

Campbell, A. N. C., Nunes, E. V., Miele, G. M., Matthews, A., Polsky, D., Ghitza, U. E., et al. (2012). Design and methodological considerations of an effectiveness trial of a computer-assisted intervention: An example from the NIDA Clinical Trials Network. *Contemporary Clinical Trials, 33,* 386–395.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.

Carroll, K. M., Ball, S. A., Jackson, R., Martino, S., Petry, N. M., Stitzer, M. L., et al. (2011). Ten take home lessons from the first ten years of the CTN and ten recommendations for the future. *American Journal of Drug and Alcohol Abuse, 37,* 275–282.

Carroll, K. M., Ball, S. A., Martino, S., Nich, C., Babuscio, T. A., Nuro, K. F., et al. (2008). Computer-assisted delivery of cognitive-behavioral therapy for addiction: A randomized trial of CBT4CBT. *American Journal of Psychiatry, 165,* 881–888.

Carroll, K. M., & Rounsaville, B. J. (2003). Bridging the gap: A hybrid model to link efficacy and effectiveness research in substance abuse treatment. *Psychiatric Services, 54,* 333–339.

Carroll, K. M., & Rounsaville, B. J. (2007). A vision of the next generation of behavioral therapies in the addictions. *Addiction, 107,* 850–869.

Coffey, C. S., Levin, B., Clark, C., Timmerman, C., Wittes, J., Gilbert, P., et al. (2012). Overview, hurdles, and future work in adaptive designs: Perspectives from a National Institute Health-funded workshop. *Clinical Trials, 9,* 671–680.

Del Boca, F. K., & Darkes, J. (2007a). Enhancing the validity and utility of randomized clinical trials in addictions treatment research: I. *Participant samples and assessment. Addiction, 102,* 1047–1056.

Del Boca, F. K., & Darkes, J. (2007b). Enhancing the validity and utility of randomized clinical trials in

addictions treatment research: II. *Participant samples and assessment. Addiction, 102,* 1194–1203.

Evins, A. E., Cather, C., Pratt, S. A., Pachas, G. N., Hoeppner, S. S., Goff, D. C., et al. (2014). Maintenance treatment with varenicline for smoking cessation in patients with schizophrenia and bipolar disorder. *Journal of the American Medical Association, 311,* 145–154.

Feaster, D. J., Mikulich-Glibertson, S., & Brincks, A. M. (2011). Modeling site effects in the design and analysis of multisite trials. *American Journal of Drug and Alcohol Abuse, 37,* 383–391.

Fernández-Montalvo, J., López-Goñi, J. J., Illescas, C., Landa, N., & Norea, I. (2008). Evaluation of a therapeutic treatment program: A long-term follow-up study in Spain. *Substance Use and Misuse, 43,* 1362–1377.

Fletcher, B. W., Tims, F. M., & Brown, B. S. (1997). Drug abuse treatment outcome study (DATOS): Treatment evaluation research in the United States. *Psychology of Addictive Behaviors, 11,* 216–229.

Flynn, P. M., & Brown, B. S. (2008). Co-occurring disorders in substance abuse treatment: Issues and prospects. *Journal of Substance Abuse Treatment, 34,* 36–47.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York, NY: Springer-Verlag.

Gerstein, D., Datta, R., Ingels, J., Johnson, R., Rasinski, K., Schildhaus, S., Talley, K., Jordan, K., Phillips, D. B., Anderson, D. W., Condelli, W. G., & Collins, J. S. (1997). *National Treatment Improvement Evaluation Survey, Final Report*, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration. Washington, DC: US Department of Health and Human Services.

Gibbons, R. D., Hedeker, D., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology, 6,* 79–107.

Gossop, M., Marsden, J., Stewart, D., & Rolfe, A. (1999). Treatment retention and 1 year outcomes for residential programmes in England. *Drug and Alcohol Dependence, 57,* 89–98.

Grossman, J., & MacKenzie, F. J. (2005). The randomized controlled trial: Gold standard, or merely standard? *Perspectives in Biology and Medicine, 48,* 516–534.

Hawkins, N. G., Sanson-Fisher, R. W., Shakeshaft, A., D'Este, C., & Green, L. W. (2007). The multiple baseline design for evaluating population-based research. *American Journal of Preventive Medicine, 33,* 162–168.

Hedden, S. L., Woolson, R. F., Carter, R. E., Palesch, Y., Upadhyaya, H. P., & Malcom, R. J. (2009). The impact of loss to follow-up on hypothesis tests of the treatment for several statistical methods in substance abuse clinical trials. *Journal of Substance Abuse Treatment, 37,* 54–63.

Hedden, S. L., Woolson, R. F., & Malcolm, R. J. (2006). Randomization in substance abuse clinical trials. *Substance Abuse Treatment, Prevention, and Policy, 1,* 6. Epub. doi:10.1186/1747-597X-1-6

Hien, D. A., Wells, E. A., Jiang, H., Suarez-Morales, L., Campbell, A. N. C., Cohen, L. R., et al. (2009). Multi-site randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. *Journal of Consulting and Clinical Psychology, 77,* 607–619.

Hser, Y.-I., Evans, E., Huang, D., Weiss, R. D., Saxon, A. J., Carroll, K. M., et al. (2015). Long-term outcomes after randomization to Buprenorphine/Naloxone versus Methadone in a multi-site trial. *Addiction, 111*(4), 695–705.

Hser, Y.-I., Saxon, A. J., Huang, D., Hasson, A., Thomas, C., Hillhouse, M., et al. (2014). Treatment retention among patients randomized to buprenorphine/naloxone compared to methadone in a multi-site trial. *Addiction, 109,* 79–87.

Hubbard, R. L., Craddock, S. G., & Anderson, J. (2003). Overview of 5-year follow-up outcomes in the Drug Abuse Treatment Outcome Studies (DATOS). *Journal of Substance Abuse Treatment, 25,* 125–134.

Humphreys, K., Horst, D., Joshi, A. A., & Finnery, J. W. (2005). Prevalence and predictors of research participant eligibility criteria in alcohol treatment outcome studies, 1970–98. *Addiction, 100,* 1249–1257.

Humpreys, K., Maisel, N. C., Blodgett, J. C., & Finney, J. W. (2013). Representativeness of patients enrolled in influential clinical trials: A comparison of substance dependence with other medical disorders. *Journal of Studies on Alcohol and Drugs, 74,* 889–893.

Jones, H. E., O'Grady, K. E., & Tuten, M. (2011). Reinforcement-based treatment improves the maternal treatment and neonatal outcomes of pregnant patients enrolled in comprehensive care treatment. *The American Journal on Addictions, 20,* 196–204.

Kahler, C. W., Metrik, J., LaChance, H. R., Ramsey, S. E., Abrams, D. B., Monti, P. M., et al. (2009). Addressing heavy drinking in smoking cessation treatment: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 76,* 852–862.

Kao, L. S., Tyson, J. E., Blakely, M. L., & Lally, K. P. (2008). Clinical research methodology I: Introduction to randomized trials. *Journal of the American College of Surgeons, 206,* 361–369.

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology, 54,* 541–549.

Karlsson, P., & Bergmark, A. (2014). Compared with what? An analysis of control group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders. *Addiction, 110,* 420–428.

Kunz, R., Vist, G., Oxman, A. D. (2008). Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews, 2*. doi:10.1002/14651858.MR000012.pub2

Ling, W., Amass, L., Shoptaw, S., Annon, J.J., Hillhouse, M., Babcock, D., Brigham, G., Harrer, J., Reid, M., Muir, J., Buchan, B., Orr, D., Woody, G., Krejci, J.,

Ziedonis, D, Buprenorphine Study Protocol Group. (2005). A multi-center randomized trial of burprenorphine-naloxone versus clonidine for opioid detoxification: Findings from the National Institute on Drug Abuse Clinical Trials Network. *Addiction, 100,* 1090–1100.

Machin, D., & Fayers, P. M. (2010). *Randomized clinical trials design, practice and reporting*. Hoboken, NJ: Wiley-Blackwell.

Marsch, L. A., Guarino, H., Asosta, M., Aponte-Melendez, Y., Cleland, C., Grabinski, M., et al. (2014). Web-based behavioral treatment for substance use disorders as a partial replacement of standard methadone maintenance treatment. *Journal of Substance Abuse Treatment, 46,* 43–51.

McGovern, M. P., Fox, T. S., Xie, H., & Drake, R. E. (2004). A survey of clinical practices and readiness to adopt evidence-based practices: Dissemination research in an addiction treatment system. *Journal of Substance Abuse Treatment, 26,* 305–312.

McMahon, R. C., Kouzekanani, K., & Malow, R. M. (1999). A comparative study of cocaine-treatment completers and dropouts. *Journal of Substance Abuse Treatment, 16,* 17–22.

McPherson, S., Barbosa-Leiker, C., Mamey, M. R., McDonell, M., Enders, C. K., & Roll, J. M. (2015). A 'missing not at *random'* (MNAR) and 'missing at random (MAR) growth model comparison with a Buprenorphine/Naloxone clinical trial. *Addiction, 110*(1), 51–58.

Melberg, H. O., & Humphreys, K. (2010). Ineligibility and refusal to participate in randomised trials of treatments for drug dependence. *Drug and Alcohol Review, 29,* 193–201.

Mercer, S. L., DeVinney, B. J., Fine, L. J., Green, L. W., & Dougherty, D. (2007). Study designs for effectiveness and translation research. *American Journal of Preventive Medicine, 33,* 139–154.

Murphy, S. A., Lynch, K. G., Oslin, D., McKay, J. R., & TenHave, T. (2007). Developing adaptive treatment strategies in substance abuse research. *Drug and Alcohol Dependence, 88*(suppl. 2), S24–S30.

Nunes, E. V., Ball, S., Booth, R., Brigham, G., Calsyn, D. A., Carroll, K., et al. (2010). Multi-site effectiveness trials of treatments for substance abuse and co-occurring problems: Have we chosen the best designs? *Journal of Substance Abuse Treatment, 38* (suppl. 1), S97–S112.

Orford, J. (2008). Asking the right questions in the right way: The need for a shift in research on psychological treatments for addiction. *Addiction, 103,* 886–892.

Pagoto, S. L., Kozak, A. T., John, P., Bodenlos, J. S., Hedeker, D., Spring, B., et al. (2009). Intention-to-treat analyses in behavioral medicine in randomized clinical trials. *International Journal of Behavioral Medicine, 16,* 316–322.

Perron, B. E., Cordova, D., Salas-Wright, C., & Vaughn, M. G. (2017). Validity: Conceptual and methodological issues in substance abuse research. In J. B. VanGeest, T. P. Johnson, & S. Alemagno (Eds.), *Research methods in the study of substance abuse* (pp. XX–XX). New York, NY: Springer.

Petri, N. M., Peirce, J. M., Stitzer, M. L., Blaine, J., Roll, J. M., Cohen, A., et al. (2005). Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs. *Archives of General Psychiatry, 62,* 1148–1156.

Pettinati, H. M., Silverman, B. L., Battisti, J. J., Forman, R., Schweizer, E., & Gastfriend, D. R. (2011). Efficacy of extended-release naltrexone in patients with relatively higher severity of alcohol dependence. *Alcoholism, Clinical and Experimental Research, 35,* 1804–1811.

Retrosen, J., Leshner, A., Tai, B., Greenlick, M., Pencer, E., Trachtenberg, R., Woody, G. (2002). National Drug Abuse Clinical Trials Network—Challenges and opportunites. In L. S. Harris (Ed.), *Problems of Drug Dependence, 2001: Proceedings of the 63rd Annual Scientific Meeting, the College on Problems or Drug Dependence, Inc.* NIDA Research Monograph 182 (pp. 12–17). Bethesda, MD: National Institute on Drug Abuse.

Roan, S. (November 10, 2008). The 30-day myth. *Los Angeles Times.* Retrieved November 24, 2014 from http://articles.latimes.com/2008/nov/10/health/he-addiction10

Rosner, A. (2002). Reality-based evidence: Prospecting for the elusive gold standard. *Dynamic Chiropractic, 20,* 14–16.

Rounsaville, B. J., Carroll, K. M., & Onken, L. S. (2001). NIDA's stage model of behavioral therapies research: Getting started and moving on from Stage I. *Clinical Psychology: Science and Practice, 8,* 133–142.

Scott, C. K. (2004). A replicable model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. *Drug and Alcohol Dependence, 74,* 21–36.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized quasi-experimental inference* (2nd ed.). New York, NY: Houghton Mifflin.

Solomon, P., Cavanaugh, M. M., & Draine, J. (2009). *Randomized controlled trials: Design and implementation for community-based psychosocial interventions (pocket guides to social work research methods)*. New York, NY: Oxford University Press.

Stout, R. L., Wirtz, P. W., Carbonari, J. P., & Del Boca, F. K. (1994). Ensuring balanced distribution of prognostic factors in treatment outcome research. *Journal of Studies on Alcohol* (suppl. 12), 70–75.

Susikida, R., Crum, R. M., Stuart, E. A., Ebnesajjad, C., & Motjabai, R. (2016). Assessing sample representativeness in randomized clinical trials: Application to the National Institute on Drug Abuse Clinical Trials Network. *Addiction, 111*, 1226–1234.

Swartz, J. A. (2012). *Substance Abuse in America: A documentary and reference guide*. Santa Barbara, CA: ABC-Clio.

Tai, B., Sparenborg, D., Liu, D., & Straus, M. (2011). The national drug abuse treatment clinical trials network: Forging a partnership between research knowledge and community practice. *Substance Abuse and Rehabilitation, 2,* 21–28.

Tiffany, S. T., Friedman, L., Greenfield, S. F., Hasin, D. S., & Jackson, R. (2012). Beyond drug use: A systematic consideration of other outcomes in evaluations of treatments for substance use disorders. *Addiction, 107,* 709–718.

Wallace, P., Struzzo, P., della Vedova, R., Tersar, C., Verbano, L., Lygidakis, H., MacGregor, R., Freemantle, N., & Scafato, E. (2013). Randomised controlled non-inferiority trial of primary care based facilitated access to an alcohol reduction website (EFAR-FVG). *Addiction Science & Clinical Practice, 8* (Suppl 1), A83.

Witkiewitz, K., Warner, K., Sully, B., Barricks, A., Stauffer, C., Thompson, B. L., et al. (2014). Randomized trial comparing mindfulness-based relapse prevention with relapse prevention for women offenders at a residential addiction treatment center. *Substance Use and Misuse, 49,* 536–546.

Wolff, N. (2000). Using randomized controlled trials to evaluate socially complex services: Problems, challenges and recommendations. *The Journal of Mental Health Policy and Economics, 3,* 97–109.

Woolf, S. H. (2008). The meaning of translational research and why it matters. *Journal of the American Medical Association, 209,* 211–213.

Zhang, Z., Friedmann, P. D., & Gerstein, D. R. (2002). Does retention matter? Treatment duration and improvement in drug use. *Addiction, 98,* 673–684.

Sampling Strategies for Substance Abuse Research

4

Joseph Gfroerer, Arthur Hughes, and Jonaki Bose

4.1 Introduction

In studies of substance abuse among a defined population of interest, it is often not feasible to collect data from every member of the population. The cost may be too high, data collection might take too long, or there may be other logistical barriers. Fortunately, most studies can be successfully carried out by collecting data from a carefully drawn sample of the population. This chapter discusses issues faced in developing strategies to draw such a sample, i.e., the sample designs, and criteria that can be used to identify optimal sampling methods for substance abuse studies. The goal of the chapter is to provide guidance for researchers to use in making decisions about study designs, by explaining the strengths and weaknesses of different options, and demonstrating how different designs have been used in actual studies. The information in this chapter should also be helpful to researchers involved in secondary analysis of data sets based on samples, because of the impact of sample designs on analytic methods and data interpretation. Though this chapter focuses on standard probability-based surveys, non-probability sampling, which generally should not involve producing prevalence estimates that are representative of a population of inference, is also discussed.

The next section includes a description of the process of determining sampling strategies for probability samples and includes illustrations of sample design principles with examples of substance abuse surveys. Section 4.3 contains more specific discussions of concerns relevant for specific types of substance abuse studies, including general population surveys and surveys of students, criminal justice populations, and treatment populations. Unique problems encountered when conducting these types of substance abuse studies are highlighted. A subsequent section discusses non-probability sampling, and a final summary section reiterates the main points of the chapter.

4.2 General Sample Design Principles for Probability Samples

Before establishing a sample design, it is critical to have an understanding of general sampling principles, practices, and terminology. This section provides an overview of some basic concepts and terminology which are relevant for

J. Gfroerer (✉) · A. Hughes · J. Bose
Substance Abuse and Mental Health Services Administration, Rockville, MD, USA
e-mail: Joe.Gfroerer@msn.com

A. Hughes
e-mail: arthugh80@gmail.com

J. Bose
e-mail: Jonaki.bose@samhsa.hhs.gov

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_4

studies using samples regardless of the topic and population of interest. We provide a brief overview of key sample design issues, and suggest that readers consult some of the excellent general sample design books listed in the references for more in-depth discussion (Kish 1965; Cochran 1977; Groves 2004; Groves et al. 2009; Thompson 2012; Lohr 2010). A short primer on designing surveys is available through the American Statistical Association (Scheuren 2004).

The sample design should reflect the goals and purpose of the study. This may seem to be a trivial and obvious statement, but in practice, this fundamental principle is sometimes overlooked as study designs evolve and eventually become finalized. Thus, it is a good idea to periodically revisit the original goals of the study to ensure the design planning continues to remain consistent with these goals. This is particularly important if the sample designers are not part of the core study team that has the clearest understanding of the purpose of the study.

Studies typically have multiple analytic goals, and each goal might be best achieved by using a different sampling plan. For example, a survey could be designed to estimate the prevalence of substance use disorder as well as the prevalence of co-morbid health conditions. Communication between the study leaders and sampling statisticians is essential in order to determine a single sampling plan that optimally addresses all of the study goals. Such a plan would consider competing priorities among the goals in conjunction with the tradeoffs in cost and data quality resulting from different design options. The benefits of communications, early and often, between analysts and survey designers cannot be overemphasized. Analysis plans should be developed and discussed with sample designers prior to the sample design phase, and adjusted once the final sampling plan is determined. These analysis plans should identify key outcome measures, as well as the kinds of analysis that are intended, including comparisons to be made between subgroups.

The use of consistent, commonly understood and accurate terminology within a team facilitates communication. Although somewhat basic, below are some brief definitions for key terms to help navigate this discussion

Target population—The entire population or universe of units to be studied.

Sample frame—A complete or nearly complete list of target population units from which a sample can be drawn. The sample frame should be as close to the target population as possible, but may differ due to practical or other constraints.

Sample—A subset of target population units that has been selected from the sample frame to provide information. Depending on how the sample is selected, it may or may not be representative of the target population.

Sample design—The process through which the target population is defined, the sampling frame is created, and the sample is drawn.

Probability sample—A sample in which every unit in the sampling frame has a known nonzero probability of selection into the sample.

Simple random sample—A sample in which every unit selected from the sampling frame has an equal probability of selection and, for a sample of size n, every possible combination of n sample units in the frame has an equal chance of being in the sample.

Stratification—The division of the sampling frame into separate groups based on characteristics of interest and then drawing samples within each of these separate groups or strata. This is done to control sample sizes within strata, which enables stratum-specific estimation or analysis, may reduce sampling error, and can help control costs.

Cluster sampling—The process in which the target population in the sampling frame is divided into groups or clusters (e.g., counties, telephone exchanges, or healthcare facilities), and a sample of these clusters is selected. Data are collected from individual members (or a subsample) of the units in each sample cluster. In surveys where data collection requires travel to the location of the sample unit, cluster sampling is commonly used to reduce the average cost per interview.

Complex sample—A sample that is not a simple random sample. A complex sample

incorporates a combination of techniques such as stratification, clustering, multiple stages of sampling, and unequal probability sampling such that direct computation of estimates of sampling errors is not feasible, usually requiring specialized software. Note that although cluster sampling and stratification can be applied in ways that give each member of the target population an approximately equal probability of selection, this would still constitute a complex sample design.

Coverage error—The difference between the target population and the sample frame. If those excluded from the sample frame are different than the target population for the measures of interest, then coverage bias may be introduced in the results.

Sampling error—The error caused by observing a sample instead of the entire population. The difference between the true population value and the estimate derived from the sample is a measure of sampling error. However, since the true population value is usually unknown, probabilistic methods have been developed to estimate sampling error.

Design effect—The ratio of the variance from the actual sample design (often a complex sample design) and the variance if the sample had been obtained via a simple random sample.

Effective sample size—The actual sample size divided by the design effect, i.e., the sample size that would produce the same variance had a simple random sample been used.

Nonresponse—Nonresponse occurs when the sampled units do not participate in the survey entirely (unit nonresponse) or when survey participants do not provide responses on certain items within the data collection instrument (item nonresponse) When the respondents differ from the nonrespondents in the measures of interest, nonresponse bias may be introduced in the study estimates.

Weighting—A process that attempts to ensure that estimates from the sampled cases represent the target population as closely as possible. Typically, a weight assigned to an individual sample unit denotes the estimated number of target population members "represented" by that sample unit. Weighting takes into account the probability of being included in the sample, and often includes components to correct for unit nonresponse and coverage error.

Once the goals of the study have been defined, the development of the sampling plan can begin. A first step is to specify a target population and a sampling frame. Often, these are different, due to practical limitations. For example, for a study of use patterns among current heroin users (the target population) in a community, one would like to have a list or registry of all heroin users in the community (sample frame) from which to select a sample of users to locate and interview. But such a registry is not likely to be available, and it would be difficult and costly to construct one that would be complete. Alternative sample frames could be developed from households (either by phone or face-to-face), the Internet, treatment facilities, booking centers, emergency rooms, or street contacts who know heroin users. In choosing which particular frame to use, considerations include the cost of obtaining the frame, the coverage error, and the amount of available information on each sample unit on the frame, which could be important in designing the sample (e.g., stratification) and in post-study analysis (e.g., in evaluating the nonresponse bias and weighting).

Depending on the goals of the study, either a cross-sectional or a longitudinal design can be employed. In a cross-sectional design, data are collected from each sample unit just one time, while longitudinal designs involve the collection of data from each sample unit at multiple predetermined points in time. In general, higher cost and logistical considerations (e.g., tracking individuals) of longitudinal designs mean that their sample sizes tend to be smaller than cross-sectional designs. But if the study goals include assessing causation, onset or the effects of changes over time on outcomes such as substance use, a longitudinal design has analytical advantages over repeated cross-sectional studies (Bachman et al. 2011). In some cases, cross-sectional designs may collect historical data by asking about past key events and behaviors, such as marital history and drug use at prior ages, to generate longitudinal data, but this

approach can be problematic due to errors in recall (Gfroerer et al. 2004).

Designing an effective sample design requires the optimization between study goals, cost, time, and other logistical constraints. Designing a sample is an analytic task that requires data on factors such as the expected values and variances of key estimates, differences in estimates and variances across potential strata, homogeneity of clusters, and response rates. The more data that can be brought into the design task and the better the quality of that data, the more efficient the design will be in meeting the study goals. Along with the choice of sampling frame, the size and structure of the sample selected from that frame are all components of the design. The structure of the sample is primarily driven by stratification and clustering. These can be determined in a systematic and thoughtful way so as to maximize the utility of the resulting data by reducing sampling error, ensuring populations of interest are represented, and by controlling data collection costs.

4.2.1 Stratification

Stratification involves classifying units in the sampling frame into separate groups, or strata, and then selecting a sample of units within each stratum. One of the goals of stratification may be to ensure certain groups have large enough samples for adequate analysis. For example, if the study goals include estimation of marijuana use by racial/ethnic groups, stratification by racial/ethnic group may be necessary to ensure there are enough American Indian/Alaska Natives or Asians in the sample. When the sample design specifies a higher selection rate in these small population groups it is referred to as "oversampling." This is an example of the need for good data for sample design purposes. In order to oversample a specific population, or in any way control sampling rates among different segments of the target population, accurate information on key classifying variables (race/ethnicity in this example) is needed for every sample unit in the frame. In the absence of that information, a costly screening procedure might be necessary to determine the race-ethnicity of each unit of the sampling frame. A lower cost approach for oversampling often employed in multistage sample designs is to utilize existing data on the distribution of the particular subpopulations of interest. For example, geographic areas known to have high concentrations of the race/ethnic groups of interest can be oversampled. Similarly, sampling rates in a survey of patients enrolled in treatment programs can be varied across types of facilities to achieve oversampling of target particular types of patients, given prior knowledge of the patient mix at each facility.

When armed with relevant data on the characteristics of the members of the target population and the key variables of interest, stratification can also be used to structure the sample to minimize cost and maximize precision (i.e., minimize sampling error). This involves applying higher sampling rates among subgroups (defined as separate strata) where the variance of the outcome of interest is higher. Consider a hypothetical study in which a random sample of 1000 hospital inpatients is to be selected for a follow-up interview asking about their alcohol use. The sample frame consists of the medical records from 50,000 admissions in five hospitals within the community being studied for a given year. Each hospital accounts for about 10,000 admissions per year, so a reasonable approach would be to randomly select 200 admissions from each hospital. However, suppose from some earlier study we know that the community has one primary alcohol inpatient treatment center, located in hospital A, and that most heavy drinkers, when they need care, are admitted to hospital A. The other four hospitals serve few heavy drinkers, with a mix of mostly moderate and light drinkers, and abstainers. Statistically, the variance of the patients' alcohol consumption is substantially greater in hospital A than in the other four hospitals. With this data, a sample design could be developed that would actually

provide a more precise estimate of consumption in the community, compared with the original design, based on the same sample size of 1000 (and applying appropriate analysis weights for estimation). This is accomplished by increasing the sampling rate in hospital A, say to 400 admissions, while selecting only 150 admissions from each of the other hospitals.

4.2.2 Cluster Sampling

A common feature of sample designs is the use of clustering. In most large-scale data collection efforts, cluster sampling is essential for cost and logistical reasons. Cluster sampling involves dividing the sample units (e.g., household members, treatment patients, students) into groups, often based on geography, and then selecting a sample of these groups or clusters. Often this results in a multistage sample design, where an initial sample of clusters, referred to as primary sampling units (PSUs) is selected and then within each of the sample clusters, a second-stage sample of units may be selected. There can be several stages.

One benefit from using cluster sampling is that a complete sampling frame listing every member of the target population is not needed. Most large-scale general population surveys would not be feasible without cluster sampling, because there is no readily available list of persons to use as a frame. Thus, in many population studies the first stage of sampling has been groups of counties formed into clusters, and then a second-stage sample of smaller geographic units such as Census tracts or block groups has been selected within the sample counties. Then a listing of addresses can be constructed within these small geographic areas. Cluster sampling not only provides a way to construct a sampling frame where none initially exists, but it also results in efficiencies in data collection, because data collection staff can be concentrated in the selected geographic areas or institutions and contact sample units more efficiently with less travel. However, there is a downside to cluster sampling. The precision of the estimates is nearly always reduced when clustering is part of the sample design. Briefly, the impact on precision is determined by the size and number of clusters in the sample, along with the intra-cluster correlation coefficient (referred to by Groves et al. 2009 as ROH, or rate of homogeneity), which is a measure of the similarity of sample elements within a cluster. A large positive ROH and a small number of clusters (i.e., requiring a larger sample selected within each cluster) will increase the variance.

4.2.3 Practical Considerations

The preceding discussion of stratification and cluster sampling is intended to provide a basic understanding of why these techniques are used in designing samples and what the consequences of their uses can be. But there are other factors that must be considered in the design of a sample. Of utmost importance is cost, which plays a major role in the design. Often a study may have a fixed budget and the goal might be to develop the best sample design within that budget. Or the goals of the study must be met in which case a different sample design options and their associated costs that meet the goals may be explored. Often, the overall sample size of a study is the main determinant of the cost.

Stratification and oversampling can be used to improve precision (i.e., reduced sampling error) for overall estimates or for particular subpopulations of interest, but there may be impacts on costs, e.g., if the cost of data collection is higher in the oversampled groups. Similarly, clustering, which is typically used to reduce costs, most often results in reduced precision. Sampling textbooks usually include a discussion of optimal allocation, in which computations can be performed to determine the best balance between cost and sampling error in a sample design, given a fixed sample size, or the sample allocation that would result in the lowest sampling error for a fixed overall survey cost. However, for large-scale studies in which complex samples are

needed, sampling theory may not provide a clear path for sample designers. Tradeoffs in cost and precision are difficult to quantify for designs that combine stratification with sampling of clusters and subsampling within clusters, especially where there are varying data collection costs across clusters and strata. This is an area that needs further research (Groves et al. 2009). A general rule of thumb is that sampling efficiency (i.e., maximum precision for a fixed sample size) is achieved when the sample allocation specifies proportionally more sample towards the portions of the frame where the outcome variables of interest are most variable and where data collection costs are lowest. This applies not only to the construction and sampling of clusters, but also to the definition of sampling strata. Readers should consult sampling books for a more in-depth discussion of these topics.[1]

Study designers must look at multiple scenarios to get a sense of the ultimate optimal sample design, and make choices as appropriate. Once again, extensive data from prior studies and accurate assumptions are required for this exercise. Furthermore, most studies have multiple goals and measures, so it is not sufficient to optimize the sample for only one measure, e.g., past year use of marijuana. Even for a study that focuses only on one topic such as marijuana use, researchers will want to measure different aspects such as attitudes, initiation, current use patterns, dependence, and treatment. Each of these marijuana-related topics quite possibly may have a different optimal sample design. To address this, the optimization could be done on some composite statistic such as the mean of a set of key outcomes. Alternatively or in addition to the composite approach, optimization could be done separately for each key measure, and a hybrid design could be developed.

Again, accurate data on the characteristics of members of the target population and sampling frame, as well as on data collection costs, are a requirement for sample design optimization. It is also important to note that optimal sample allocations (e.g., those providing a sample size that minimizes variance for a fixed cost) are usually robust, meaning that fairly notable deviations from the optimal sample size will not significantly impact the variance. For example, if a selected sample size is 20% smaller than the optimal sample size, the increase in variance is only about 4% (Cochran 1977, page 116).

There are other factors to consider in developing the sample design. For any study, the sample design is but one component of the overall study design. The optimization procedures that strike a balance between sampling error and costs described above can assist researchers in getting accurate data at minimal cost. However, to best achieve the study goals, study designers should also consider the sources and impact of nonsampling errors, and account for them in a more comprehensive optimization. This approach has been referred to as total survey design (TSD). Nonsampling errors include frame errors, errors due to nonresponse of sample elements, item nonresponse, measurement error such as respondents providing incorrect information, and processing error, such as mistakes in coding, editing, weighting, or tabulating data. If data on the impact of these types of nonsampling error exist, they should be incorporated into optimization analyses which might potentially affect sample design decisions. For example, studies have shown that youths are more likely to answer survey questions about their drug use honestly if the survey procedures provide more privacy in the interview (Gfroerer et al. 2012, SAMHSA 2012). Therefore study designers may want to consider the cost and operational aspects of using a self-administered questionnaire versus an interviewer administered instrument, and the added cost and impact of enhancing the confidentiality protections and promises made to respondents. Allocating more resources towards reducing nonsampling error might require a reduction in sample size or some other sample design modification to keep within the study budget, but may result in higher quality data overall.

[1]See for example, Kish (1965), Cochran (1977), Groves (2004), and Lohr (2010). Also see Chaps. 1 and 2 in Korn and Graubard (1999) for an excellent primer on sampling methodologies common in health surveys.

4.3 Probability Sampling for Specific Types of Substance Use Studies

While general sampling theory is applicable to surveys of any topic area, there are particular concerns that arise in designing samples for substance abuse surveys. Some of these are described below.

4.3.1 General Population Surveys

General population surveys can be implemented through face-to-face surveys using area probability sampling, telephone data collection, by mail, via the web, or through some combination of these modalities. Because of the potential for low cost and fast turnaround, there is great interest in collecting substance use data via the internet. Current internet panels cannot be classified as probability samples due to factors such as non-probabilistic recruitment into the panel and low response rates. More research is needed to refine the sampling methods for Internet panels to produce probability samples that can produce unbiased population estimates. Telephone surveys have often been used to collect data on substance use, but there are concerns about nonresponse bias due to very low participation rates in recent years. For example, the 2011 California Behavioral Risk Factor Surveillance System (BRFSS) combined landline telephone and cell phone response rate was 35.4%. The 2011 BRFSS response rates ranged from 33.8 to 64.1% across states, with a median response rate of 49.7% (CDC 2013a). The 2011–2012 California Health Interview Survey (CHIS), also conducted by both landline telephone and cell phone, achieved a response rate of 35.1% (computed using the BRFSS methodology) (UCLA 2014). Coverage has also become a major concern due to the increasing prevalence of U.S. households with no landline (Blumberg and Luke 2013) who are missed using traditional telephone survey sampling frame construction procedures. To address the potential for coverage bias, in 2011 the BRFSS added a cell phone component to the sample frame. The 2013 BRFSS and 2011–2012 CHIS landline response rates were 49.6 and 39.5% respectively while the cell phone response rates were 27.9 and 32.1% respectively. While coverage of households improved with the addition of cell phone sample, nonresponse bias remains a concern. Despite the low rate for cell phone participation, it was similar to or higher than response rates in other selected studies where cell phones are used.

The largest substance use study conducted in the U.S. is the National Survey on Drug Use and Health (NSDUH), which collects data in face-to-face interviews. The high cost of face-to-face data collection (on average, several hundred dollars per completed interview) is a limitation for most studies. However, response rates tend to be higher with face-to-face surveys, and the face-to-face setting generally permits a more in-depth interview. The design of the NSDUH sample incorporates features described in the previous section, applied in the context of the specific substance use indicators of interest to the Federal Government. Like most large national surveys, NSDUH employs a multistage area sample. The target population is the civilian noninstitutionalized population of the United States, aged 12 and older. Most national surveys start with a first-stage selection of PSUs that consist of counties or groups of counties, and a second-stage sample of smaller geographic areas. Typically, this type of design will result in some states having very few or even zero PSUs in the sample, which is not a problem and standard for studies only requiring national estimates. But NSDUH requires estimates for every state. Thus, to ensure sample sizes are adequate in every state, the first-stage selection (for 2014) is 6000 Census tracts. These tracts are selected from a frame that is divided into 750 strata, formed by aggregating adjacent tracts such that the population of all strata within a given state is approximately equal. The rationale behind equalizing the population within state and stratum is to equalize interviewer workload although, in practice, workloads may vary if some strata are more difficult to obtain respondent cooperation than others. In each sample

tract, a subsample of one small cluster, called a segment, is selected. In these segments, which typically consist of a minimum of 100–250 housing units depending on the state and urbanicity, field staff constructs listings of addresses, which serve as the sampling frame for the next stage of selection which are housing units. Selected housing units are contacted, and individual members are listed in a screening interview so that the final stage of sampling can be implemented: the selection of zero, one, or two persons in the housing unit. At this phase of sampling, it is necessary to obtain some basic demographic information on housing unit members to enable stratification by age, and also for use in sample weighting. The screening interview is necessary to obtain an oversample of young people (e.g., 25% of the sample is age 12–17 and 25% of the sample is age 18–25) and results in a need to successfully contact and screen approximately 120,000 housing units in order to locate and select the target sample of 67,500 completed interviews with the required age distribution.

The construction of the listing of housing units in segments is labor-intensive and expensive. In recent years, evaluation studies of an alternative sampling frame for the NSDUH based on a computerized list of addresses were conducted. This list frame is maintained and updated by the U.S. Postal Service, is increasingly becoming a viable alternative to traditional listing approaches, and has the potential to significantly reduce costs. Several limitations currently exist for populations living in rural areas and group quarters such as college dormitories. In rural areas, there are a considerable number of residences that do not have a city-style address (i.e., house number street, city, state, zip code) making it difficult if not impossible for an interviewer to locate (Iannacchione 2012). For group quarters, there is uncertainty about where to obtain an accurate source that can be used to identify segments with high concentrations of group quarters units (e.g., college dormitory populations). However, this limitation may justify the development and use of a hybrid frame where field enumeration is used in area segments where the ABS coverage is low.

Regardless of which method is used to select housing units, survey designers have flexibility in specifying how individuals within the household are selected. An optimal protocol would account for current and future data needs, as well as cost, sampling error, and data quality considerations. Data collection costs can usually be reduced by selecting more respondents per household, which would result in fewer households needing to be contacted overall, assuming a fixed target number of completed interviews. But since households constitute clusters, a larger sample size within households would tend to result in increased sampling error if there is a positive intra-cluster correlation coefficient. In other words, if substance use behaviors tend to be similar among household members, then the most precise prevalence estimates may be achieved by selecting only one person per household. Numerous studies have shown a strong positive correlation between household members in their tobacco, alcohol, and illicit drug use (Gfroerer 1987; Kandel and Griesler 2001; and Ashley et al. 2008). In terms of sampling error, one should also consider the impact of selecting only one respondent in very large households, which could lead to large variations in sampling weights, and corresponding increases in sampling error. Another factor to consider is the burden and interview time for families when several family members are interviewed in a survey, as well as the potential effect on data quality if interviews are done on different occasions and the sampled household members discuss their responses in the interim. Nevertheless, data needs may overshadow many of these concerns. There is often great value in or even a specific need for having data on multiple household members, to address family and household issues such as marital relations and parent–child characteristics. For more discussion on within-household sampling strategies, see Brewer (1963), Chromy and Penne (2002), Cochran (1977), Sampford (1967), and Iannacchione and Shook-Sa (2013).

In the design of the sample and specification of the target population for a survey of the general population, it is important to have clear rules

related to residency within the household to determine eligibility, in order to minimize double counting or exclusion of individuals from the sample frame. This is particularly important for substance abuse studies, where some heavy drug users may have less stable living arrangements and therefore might be less likely to be included when enumerating the persons living in the housing unit. Young adults, who typically have high rates of use for many substances, tend to be more mobile and in transitional situations. The college population can be difficult to capture cleanly, with shifts between living with parents and in dormitories and other near-campus housing. It is not entirely clear which is the best approach for sampling college students in a general population survey. They could be sampled as a regular part of the area sampling, or alternatively they could be sampled through a separate sample from a list of colleges and enrolled students within those colleges. The focus of the data collection effort may guide the decisions regarding the optimal sampling frame and methodology.

Surveys on substance use in the general population are challenging because of the sensitivity of the topic for many respondents who may not want to admit to illegal or socially undesirable behaviors. The sampling and data collection protocols should be designed to minimize any negative impact that could occur due to public or individual concerns about the topic under study. Any appearance that the study is targeting drug abusers might raise suspicions among potential respondents, leading to nonresponse or inaccurate reporting by respondents. Field interviewers and other survey staff with responsibility for explaining the study to potential respondents need to be well trained to be able to explain how the sample is selected. Some sampled individuals may say "nobody in this house uses drugs, so try going next door" or others may think "I wonder if they selected me because they know or suspect that I use marijuana." Procedures and communications with the community should emphasize the random selection and confidential nature of the data collection. Although cluster sampling is essential in most large-scale general population surveys, it may be preferable to be sure that selected households have some distance between them, so that neighbors do not talk among themselves during the data collection period. More broadly, consideration should be given to whether or not a public notice of upcoming data collection would be helpful or not. Notifying local law enforcement and public health agencies prior to data collection is generally helpful, if not essential, because some sampled individuals who are suspicious will undoubtedly call these local officials to verify the legitimacy of the study. When children are interviewed in a general population survey, extra care should be taken in within-household sample selection and interview administration to ensure parental acceptance and privacy for the children during the interviews.

In addition to skepticism about response validity due to under reporting or not reporting use at all in general population surveys of substance use, a common concern among research-ers is the coverage of addicts and heavy substance abusers. A priori, many researchers assume that the most problematic substance abusers simply will not be captured in samples of the general population, because they are not in stable living situations conducive to sampling. They may be homeless, transients, in jails or treatment facilities, or otherwise not easily located. This perception is compounded by the unexpectedly low prevalence estimates produced by surveys for populations such as heroin addicts, heavy cocaine users, and injection drug users. More research is needed to resolve these concerns. For example, while it is recognized that general population surveys such as NSDUH produce undercounts of these important indicators, the magnitude of the undercount is unknown. The source of the undercount is also a key question—are the hard-core drug users missing from the frame, or are they captured by the frame but not in the sample because they refuse to participate or cannot be found? Or, are they included in the interviewed sample but do not report their high level of drug use? These are important areas needing exploration. Some studies have indicated that NSDUH covers criminal justice populations (arrestees, and those on probation or parole) and treatment populations

well, but by definition the sample excludes incarcerated or those in long-term residential treatment, as well as homeless not living in shelters (Lattimore et al. 2014; Feucht and Gfroerer 2011; Batts et al. 2014). There are numerous research studies that have tried alternative sampling and estimation methods such as snowball sampling, nominative estimation, network sampling, and capture recapture to study heavy drug users (Gfroerer and Kennet 2015; Wagner and Lee 2015).Such non-probabilistic methods are described later in this chapter in Sect. 4.4.

4.3.2 Student Surveys

Many studies of substance abuse focus on young people. Use of most substances begins at a young age, and prevalence rates for many substances are highest among young adults. To efficiently sample young people, high schools and colleges are often targeted for data collection among students. Collecting data in a school setting results in a target population consisting of students attending schools, not young people overall. School-based surveys generally do not include dropouts. In addition, even though in theory the sampling frame includes all students on the day of the interview, some students may be absent, and therefore students with higher rates of absenteeism are less likely to be included in school-based surveys. In terms of sampling and field procedures, it is much simpler to exclude dropouts and absentees, although they are both known to have elevated rates of substance abuse (SAMHSA 2012). However, the inclusion of these groups would require additional steps, such as a creating a separate frame or scheduling follow-up visits. A survey design team makes these decisions based on a multitude of factors, but ultimately it is important to be very clear about the population covered in reporting the results.

Returning to the topic of sampling schools, while it may seem a simple task to select a random sample of schools and students, there are a few considerations. Large-scale student surveys employ multistage sample designs to control costs and variances. For example, the Monitoring the Future study (MTF) uses a three-stage sample design to generate a nationally representative sample of about 13,000–16,000 high school seniors each year. Similar to household surveys, a first-stage cluster sample selects PSUs that are large-sized counties (in terms of population) or groups of smaller, adjacent counties. In the second stage of sampling, 120–150 schools (also clusters) are selected from lists of schools within each PSU. The final stage of sampling consists of randomly selecting entire classrooms (a third level of clusters) in each chosen school to capture up to about 350 students per school. Similar sample designs are used for the separate 8th and 10th grade samples. The MTF design also includes stratification and differential sampling rates. Full details on the design are available (see Bachman et al. 2011, 2014). The Youth Risk Behavior Survey (YRBS) employs a similar sample design, although the survey is conducted biennially, and samples about 15,000 9th–12th graders in about 150 schools (CDC 2013b).

The impact of clustering on sampling error for substance abuse measures is a key consideration in designing school-based student surveys. It is likely a principal reason for the high design effects found in many school surveys. To illustrate, consider a student survey such as YRBS or MTF. At the second stage of sampling in these studies, schools are selected. If for example, it is known prior to the study that the rate of cigarette use varies widely across schools (perhaps there are some schools where most students smoke, while in other schools very few students smoke), then intra-cluster correlation (ROH) at the school level would be large, because students within a given school tend to have a similar likelihood of smoking. Knowing that such a high ROH is found for student cigarette use, it would be desirable (from a precision standpoint) for the sample design to specify a large sample of schools, with relatively small numbers of students selected in each school. If ROH is negligible (i.e., the rate of smoking is about the same in every school), then there would be little loss in precision by having a fewer number of schools

and more students selected per school, which may reduce data collection costs. The MTF, which collects data from about 45,000 8th, 10th, and 12th graders in 400 schools each year, reports design effects as high as 8.1 for past month marijuana use and 5.2 for past month cigarette use among 12th graders. The YRBS design effects were 4.2 and 4.6, respectively, for 12th graders for these two measures in 2013 (Kann et al. 2014).

Other considerations in designing school surveys are the difficulties in obtaining cooperation from school districts and individual schools and whether survey designers decide to require implicit or explicit parental approval. School officials may be concerned about how the study results may be used to portray drug use in their area. Or they may be concerned about parental objections, or simply about the class time taken for survey administration. During the 1980s, MTF achieved a school response rate of around 70% in most years, but by 2002 participation had declined to about 50%. An incentive payment to schools was started in 2003, leading to a boost in the response rate to over 60%, but by 2013 participation had declined again to 54%. YRBS school response rates have been about 80%.

4.3.3 Other Types of Sampling for Substance Abuse Studies

Household and school-based samples are useful for tracking overall prevalence among broad sectors of the population and covering a wide range of substances, including tobacco and alcohol. However, some studies involve special populations of interest related to substance use such as substance abuse treatment programs and the clients they serve; criminal justice populations, such as arrestees, prison and jail inmates, and persons on probation or parole; and general healthcare facilities and their patients. These types of studies may or may not have a first-stage area sample, but generally at some point in the sampling protocol there is likely to be some contact with program administrators and the use of some kind of list of enrolled patients/inmates or a file of admission records from which a sample of persons or visits can be selected. As is the case with school surveys, obtaining cooperation from program officials is a key concern for the success of the study, so a sample design that is acceptable to administrators is helpful. In some cases, because of confidentiality concerns, program officials may prefer that sampling of patients/records be done by program staff.

4.4 Non-probability Sampling for Substance Abuse Studies

Concerns about coverage and nonresponse coupled with rising costs have led some survey researchers to consider whether non-probability sampling methods might be an acceptable alternative, at least under some conditions.

Non-probability sampling has recently become more prevalent as online data collection methods have been developed. Often, the sample used for online research is an "opt-in panel" of individuals who have been recruited in advance and agreed to participate in ad hoc surveys. The sampling approaches used with these panels vary substantially. Other types of studies that use non-probability sampling include case-control studies, clinical trials, evaluation research designs, and intercept surveys (Baker et al. 2013). Many studies employ a hybrid design, where a multistage sampling plan is used, with convenience sampling at one stage and a controlled random selection done at the final stage. Often, documentation of these studies highlights the random component and downplays the nonrandom sampling, making the study appear more valid than it truly is.

The main concern with non-probability samples is that population estimates may be highly dependent on model assumptions. The generalizability of the results based on non-probability samples depends on the appropriateness of the assumptions underlying the model and how deviations from those assumptions affect the specific estimates. When the model assumptions

are reasonably good then non-probability estimates may be accurate (Valliant and Dever 2011). However, when they are not, then the sample may exhibit selection bias where, for example, only persons who are easily approachable are recruited, resulting in a nonrepresentative sample of the target population. A famous example of a survey with significant selection bias was a telephone survey conducted during the Truman-Dewey U.S. presidential race in 1948. The results published in major newspapers stated that Dewey was projected to win by a landslide. However, what the surveyors did not realize beforehand was that low-income households were underrepresented in the sampling frame due to the high cost of owning a telephone and most importantly, people with low incomes tended to support Truman. In later years a combination of probability (random selection of polling places or random digit dialing) and non-probability samples (e.g., selecting every *k*th person at a polling place on a nonrandom basis) has been used successfully in electoral polling.

Sampling hard-to-reach or hidden populations (e.g., injection drug users, the homeless, etc.) can be challenging as persons of interest in a study and may only represent a very small segment of the total population. This makes the use of standard probability sampling methods difficult. To counter this problem, researchers have used methods such as chain-referral sampling (e.g., snowball sampling, network sampling, respondent-driven sampling), venue sampling, and adaptive cluster sampling to find persons who otherwise may be missing or hard to reach in more traditional surveys. Snowball sampling (Goodman 1961) works on the assumption that persons found to be in the population of interest, know other similar persons. So, for example, once injection drug users are located by some means, they are asked to identify and locate other injection drug users and those persons are also included in the sample. The process can continue with those persons identifying injection drug users, and so on. While this method has the potential to produce a large sample, the generalizability of the results may be based on strong modeling assumptions which may not be verifiable. Nevertheless, these methods could be useful as a means for providing information as long as the limitations are noted. Network sampling is similar in that information is collected about respondents and those respondents are asked about injection drug use by their friends or siblings. In this case, multiplicities are assigned to the respondent based on the number of persons he/she identifies as being an injection drug user. These methods can exhibit biases of unknown size and direction due to the uncertainly of the recruitment process at each stage leading to results that may not be generalizable to the population of interest.

In an effort to overcome these issues, Heckathorn (1997, 2002) developed respondent-driven sampling (RDS) that expands on the network sampling concept in two ways. First, respondents recruited at the prior stage select or 'drive' the selection of subjects at the next stage through the use of incentives to improve coverage. The second part employs a statistical model of the recruitment process that includes weighting the data to adjust for under or over coverage of the sample obtained in a nonrandom manner. Unlike weights produced from random sampling (i.e., based on a priori selection probabilities), RDS weights are produced after the cases are in the sample are identified and the size of the network and extent of homophily (the tendency of individuals to associate with other similar individuals, a characteristic often found in chain-referral samples) are taken into account. This model is based on a Markov process using transition probabilities to estimate the probability that a person recruited at a particular stage will recruit a person with a particular characteristic at the next stage. Assuming the model is correct, Heckathorn reports that RDS can provide unbiased estimates along with measures of precision and provides a website containing software to generate weights and analyze RDS data (see http://www.respondentdrivensampling.org/).

Venue sampling (Muhib et al. 2001), also known as location sampling, involves the identification and location of places or venues where the population of interest congregates (bars, night clubs, parks, etc.). These methods provide an

alternative to traditional and nontraditional sampling methods and are designed to recruit respondents in the target population in places and at times where they are expected to be found. Here, time and location is used to define the primary sampling units (PSUs). For example, a social club in the city that opens on specific days and hours during the week. The construction of the frame of PSUs is usually based on information gathered in the community through word of mouth, local magazines, etc. These units are then randomly selected, followed by sample of persons within the PSUs. Limitations of this approach are the assumption that the sampling frame is relativity complete and that persons visiting these venues on a regular basis are not different from those who attend less frequently.

Another approach is the use of adaptive cluster sampling where the rare population under study is assumed to be clustered. For example, an initial probability sample of clusters (e.g., PSUs) is selected and the number of sampled persons identified as injection drug users are obtained. If the number exceeds a predetermined value, then adjacent PSUs are included. A stopping rule is used and an estimation procedure employed to account for the sampling to produce unbiased estimates (Thompson 2012).

Banta-Green et al. (2009) used convenience samples of wastewater in municipalities in Oregon to conduct chemical analysis of drug metabolites and found the level of cocaine, methamphetamine, and MDMA use to be consistent with patterns found in epidemiological studies. While information such as incidence and prevalence cannot be obtained by this approach, this study provides evidence of the utility of wastewater-derived data for spatial analyses by improving the measurement of the level and distribution of a various illegal drugs in various communities and population centers in the state.

In other more complex non-probability surveys that measure many different phenomena, there is less evidence of their accuracy. Much more research needs to be done to determine when non-probability samples are appropriate, and how best to design them. Research is needed on how to quantify the quality of estimates from non-probability samples. Established sampling theory does not strictly apply, but it seems likely that many of the methods discussed above, such as stratification, clustering, and optimization, could be applicable to non-probability sample designs.

4.5 Estimation Based on the Sample Design

From the above discussion, it can be seen that sample designs can range from convenience samples and very simple probability designs, such as selecting every nth unit from a list of all sample units (referred to as systematic sampling) or randomly selecting any possible set of n units from a list of all sample units (simple random sample), to a multistage, stratified cluster sample with varying cluster sizes and sampling rates (complex sample design). Regardless of the sample design used, a critical principle to adhere to in any study is that the analysis of results should account for the sample design. Whenever differential sampling rates have been used, each responding case should have an associated analysis weight which reflects the sampling probability, as well as potential coverage and nonresponse adjustments. These weights should be used during analyses of the data and for any inferences made from the data. With probability sampling, in which every sample element on the frame has a known probability of selection, a base weight would typically be designated for each sample unit to be the inverse of the probability of selection for that unit. Further adjustments for nonresponse, poststratification, or coverage may also be applied to each base weight. Probability sampling allows inferences to be made reflecting the target population, and the data sets generated using probability sampling are often referred to as "representative samples." But actually these data sets are only representative of the target population if the appropriate weights are applied when producing estimates or inferences and not otherwise.

Consideration of the sample design is also necessary in analyses involving the estimation of variances and conducting significance tests. With

complex sample designs, standard formulas for computing variances cannot be used, so specialized software that takes into account sample stratification and clustering may be required to correctly estimate sampling error. This software can also be used to estimate design effects, effective n, and ROH for different population subgroups and measures, and these data can be useful in developing sample designs for other surveys, or for making adjustments to the sample of an ongoing survey. Without these correctly computed standard errors, confidence intervals and results from significance tests are likely to be inaccurate.

4.6 Conclusion

This chapter covers some of the key sampling-related concerns currently faced by substance abuse researchers. It is not possible to completely address all of the many aspects of sampling in one chapter, but through the references provided, readers can get a more comprehensive understanding of sample design for substance abuse research. While there is a multitude of sampling concerns and options that arise in substance abuse research, it may be helpful as studies are planned and carried out to keep a few basic principles in mind. Here are some recommended guidelines to follow regarding sample design for any substance abuse study:

1. Clearly state the study objectives and design the sample to achieve those objectives.
2. Communication between sample designers and other study staff is critical, including having a common understanding of key terms and aspects of sampling.
3. Use established sampling theory and search for data from prior substance abuse studies to help specify the stratification, cluster sampling, and other sample design features.
4. The sample design should consider the unique problems associated with collecting substance abuse data, such as coverage of hard-to-reach populations of drug users, stigma and privacy/confidentiality concerns, and response validity.
5. The analysis of results must account for the sample design.
6. Inferences should refer to the target population represented in the sample frame, relying on sampling theory for probability samples and stated assumptions for non-probability samples.
7. Provide complete documentation of the sample design and implementation, including coverage and response rates, estimation methods, and analysis assumptions.

Readers should also be aware that new sampling and data collection issues will continue to emerge, as new technologies and communication methods arise and evolve that could be used in sampling. Research on the use of social media and "Big Data" in statistical studies is in progress, and undoubtedly sampling will be part of that research. Regardless of what the future holds, it seems likely that these basic guidelines for sampling in substance abuse research studies will remain relevant.

References

Ashley, O. S., Penne, M. A., Loomis, K. M., Kan, M., Bauman, K. E., Aldridge, M., et al. (2008). Moderation of the association between parent and adolescent cigarette smoking by selected sociodemographic variables. *Addictive Behaviors, 33*(9), 1227–1230.

Bachman, J. G., Johnston, L. D., O'Malley, P. M., & Schulenberg, J. E. (2011). *The monitoring the future project after thirty-seven years: Design and procedures.* Monitoring the Future Occasional Paper No. 76. Ann Arbor, MI: Institute for Social Research. Available at: http://www.monitoringthefuture.org/pubs/occpapers/mtf-occ76.pdf

Bachman, J. G., Johnston, L. D., & O'Malley, P. M. (2014). *Monitoring the future: Questionnaire responses from the nation's high school seniors, 2012.* Ann Arbor, MI: Institute for Social Research. Available at: http://www.monitoringthefuture.org/datavolumes/2012/2012dv.pdf

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., et al. (2013). *Report of*

the aapor task force on non-probability sampling. Oakbrook Terrace, IL: American Association for Public Opinion Research. Available at: https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf

Banta-Green, C. J., Field, J. A., Chiaia, A. C., Sudakin, D. L., Power, L., & De Montigny, L. (2009). The spatial epidemiology of cocaine, methamphetamine and 3,4-methylenedioxy-methamphetamine (MDMA) use: A demonstration using a population measure of community drug load derived from municipal wastewater. *Addiction, 104,* 1874–1880.

Batts, K., Pemberton, M., Bose, J., Weimer, B., Henderson, L., Penne, M., et al. (2014). *Comparing and evaluating substance use treatment utilization estimates from the national survey on drug use and health and other data sources*. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Available at: http://media.samhsa.gov/data/2K14/NSDUHDR/NSDUH-DR-Task2SubUseTx-2014.pdf

Blumberg S. J., & Luke J. V. (2013). *Wireless substitution: early release of estimates from the national health interview survey, July–December 2012*. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/nhis.htm

Brewer, K. R. W. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics., 5,* 5–13.

Centers for Disease Control and Prevention. (2013a). Methodology of the youth risk behavior surveillance system-2013. *Morbidity and Mortality Weekly Report, 62,* 1.

Centers for Disease Control and Prevention (2013b). *Behavioral risk factor surveillance system—2011: Summary data quality report*. Atlanta, GA: U.S. Centers for Disease Control and Prevention. Available at: http://www.cdc.gov/brfss/annual_data/2013/pdf/2013_DQR.pdf

Chromy, J. R., & Penne, M. A. (2002). Pair sampling in household surveys. In *Proceedings of the Survey Research Methods Section, American Statistical Association*. New York, NY: American Statistical Association. Available at: https://www.amstat.org/sections/SRMS/Proceedings/y2002/Files/JSM2002–001055.pdf

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York, NY: Wiley.

Feucht, T. E., & Gfroerer, J. (2011). *Mental and substance use disorders among adult men on probation or parole: Some success against a persistent challenge, SAMHSA data review*. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Available at: http://media.samhsa.gov/data/2k11/MentalDisorders/MentalDisorders.htm.

Gfroerer, J., & Kennet, J. (2015). Collecting survey data on sensitive topics: Substance use. In T. P. Johnson (Ed.), *Handbook of health survey methods* (pp. 447–472). Hoboken: Wiley.

Gfroerer, J., Bose, J., Kroutil, L., Lopez, M., & Kann, L. (2012). Methodological considerations in estimating adolescent substance use. In *Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association. Available at: http://www.amstat.org/sections/SRMS/Proceedings/y2012/Files/304796_73741.pdf

Gfroerer, J., Hughes, A., Chromy, J., Heller, D., & Packer, L. (2004). Estimating trends in substance use based on reports of prior use in a cross-sectional survey. In S. B. Cohen & J. M. Lepkowski (Eds.), *Proceedings of the eighth conference on health survey research methods* (DHHS Publication No. PHS 04–1013, pp. 29–34). Hyattsville, MD: National Center for Health Statistics, 2004. Available at: http://www.cdc.gov/nchs/data/hsrmc/hsrmc_8th_proceedings_2004.pdf

Gfroerer, J. (1987). Correlation between drug use by teenagers and drug use by older family members. *American Journal of Drug and Alcohol Abuse, 13*(1–2), 95–108.

Goodman, L. A. (1961). Snowball sampling. *Annals of Mathematical Statistics, 32*(1), 148–170.

Groves, R. M. (2004). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). New York: Wileys.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*(2), 174–199.

Heckathorn, D. D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems, 49*(1), 11–34.

Iannacchione, V. G., McMichael, J. P., Shook-Sa, B. E., & Morton, K. B. (2012). *A proposed hybrid sampling frame for the national survey on drug use and health: Final Report*. Research Triangle Park, NC: RTI International. Available at: http://www.samhsa.gov/data/NSDUH/NSDUHMethodsRptHybrid2012.pdf

Iannacchione, V. G., & Shook-Sa, B. E. (2013). *Evaluating the effect of within- household subsampling on the precision of crime victimization rates*. (RTI Press publication No. MR-0025-1307.) Research Triangle Park, NC: RTI Press. Available at: www.rti.org/pubs/mr-0025-1307-iannacchione.pdf

Kandel, D. B., Griesler, P. C., Lee, G., Davies, M., & Schaffran, C. (2001). Parental influences on adolescent Marijuana use and the baby boom generation: Findings from the 1979–1996 national household surveys on drug abuse, analytic series A-13, DHHS Publication No. SMA 01-3531, Rockville, MD: Substance Abuse and Mental Health Services Administration.

Available at: http://media.samhsa.gov/data/NHSDA/BabyBoom/cover.htm

Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., et al. (2014). Youth risk behavior surveillance—United States, 2013, *Morbidity and Mortality Weekly Reports*, 63(4).

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Korn, E. L., & Graubard, B. I. (1999). *Analysis of health surveys*. New York: Wiley.

Lattimore, P. K., Steffey, D. M, Gfroerer, J., Bose, J., Pemberton, M. R., & Penne, M. A. (2014). *Arrestee substance use: comparison of estimates from the national survey on drug use and health and the arrestee drug abuse monitoring program*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at: http://media.samhsa.gov/data/2K14/NSDUHDRADAM/NSDUH-DR-ADAM-2014.pdf

Lohr, S. L. (2010). *Sampling: Design and Analysis* (2nd ed.). Boston: Books/Cole.

Muhib, F. B., Lin, L. S., Stueve, A., Miller, R. L., Ford, W. L., Johnson, W. D., et al. (2001). A venue-based method for sampling hard-to-reach population. *Public Health Reports, 116*(Suppl 1), 216–222.

Substance Abuse and Mental Health Services Administration. (2012). *Comparing and evaluating youth substance use estimates from the National Survey on Drug Use and Health and other surveys*, HHS Publication No. SMA 12-4727, Methodology Series M-9. Rockville, MD. Available at: http://media.samhsa.gov/data/NSDUH-M9-Youth_2012.pdf

Sampford, M. R. (1967). On Sampling without replacement with unequal probabilities of selection. *Biometrika, 54*(3–4), 499–513.

Scheuren, F. (2004). *What is a Survey*? Washington, D.C.: American Statistical Association. Available at: http://www.amstat.org/sections/srms/pamphlet.pdf

Thompson, S. K. (2012). *Sampling* (3rd ed.). New York: Wiley.

Valliant, R., & Dever, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research, 40*(1), 105–137.

UCLA Center for Health Policy Research (2014). California Health Interview Survey. *CHIS 2011–2012 Methodology Series: Report 4 – Response Rates*. Los Angeles, CA: UCLA Center for Health Policy Research. Available at: http://healthpolicy.ucla.edu/chis/design/Documents/chis2011-2012-method-2_2014-02-21.pdf. Accessed January 22, 2015.

Wagner, J., & Lee, S. (2015). Sampling rare populations. In T. P. Johnson (Ed.), *Handbook of health survey methods* (pp. 77–106). Hoboken: Wiley.

Common Statistical Methods for Primary and Secondary Analysis in Substance Abuse Research

5

Adam King, Libo Li, and Yih-Ing Hser

5.1 Introduction—Primary and Secondary Analyses

The *primary analysis* of a research study's data is the analysis conducted by the researchers who designed and executed the study, in order to address the original questions, hypotheses, and goals of that study (Glass 1976). A *secondary analysis* of the study's data is any other analysis that is either conducted by a different group of researchers, or conducted to address a different research question. Typically, secondary analyses are done by different researchers, and for different purposes, than the original primary analysis.

The use of existing datasets to examine new research questions is a common practice across a variety of disciplines, including substance abuse research. Surveys, clinical trials, cohort studies, and/or administrative/medical records, used separately or in combination, can provide information on a variety of topics related to, for example, improving the availability, accessibility, delivery, quality, effectiveness, cost-effectiveness, and outcomes of substance abuse treatment and prevention services. Analyzing existing datasets to address these new research questions or methods can maximize utility of the data already collected, given that most large-scale data collection is very costly or takes a long time to achieve. As in any research effort, it is critical to develop a conceptual framework that guides the research questions and organizes the measures/variables/constructs and relationships among them in a logical manner. Even though secondary analyses can, and often do, involve exploratory analyses, the selection of analytic methods depends on the research question to be addressed and the nature of available measures, which are then preferably guided by a strong conceptual framework.

There are, however, disadvantages to performing a secondary analysis instead of collecting new or original data. The main difficulty is finding a preexisting dataset that can directly address your research questions. One common reason that a dataset may not be adequate for secondary analysis is that the population of interest may differ from the population the dataset was sampled from; thus, conclusions drawn from an analysis of the secondary data may not be applicable to the population of interest to the secondary analyst (Smith et al. 2011). Moreover, even if the population is appropriate, variables necessary for addressing research hypotheses may not have been collected if they were not deemed necessary for addressing the hypotheses of the original study. Finally, the

A. King · L. Li · Y.-I. Hser
University of California, 11075 Santa Monica Blvd., Suite 200, Los Angeles, CA 90025, USA
e-mail: yhser@ucla.edu

A. King (✉)
Department of Mathematics and Statistics, California State Polytechnic University, 3801 West Temple Ave., Pomona, CA 91768, USA
e-mail: king@cpp.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_5

dataset may contain missing values or errors, which might be difficult to address without detailed knowledge of how the data collection process may have produced these values; such information may or may not be present in the available study documentation and/or metadata.

5.2 Study Design

This chapter begins with a brief overview of study design aspects relevant to primary and secondary analysis of substance abuse data. The first subsection focuses on topics necessary for designing a new study, while the second subsection discusses elements of study design that must be understood when performing both primary and secondary analysis.

5.2.1 Selecting a Primary Research Hypothesis and Outcome Measure

Before designing and conducting a research study, it is first necessary to formulate a *clear, specific research question*. Oftentimes, researchers begin with a research *topic*, such as the relationship between opioid replacement therapy and drug use outcomes. They must then narrow this topic down to a specific hypothesis that can be assessed by analyzing collected data, such as "heroin addicts actively taking methadone use heroin fewer days per month than subjects not on replacement therapy." Next, the researcher must determine whether this question can be assessed by data already collected from another study before investing the time and resources into designing and conducting a new study.

Once the research hypothesis has been settled upon, and no data is available to address this question in the population of interest, the organizations sponsoring or regulating the research process will usually require that this hypothesis be designated in advance as the *primary research hypothesis* of interest, from which any other ancillary questions must be clearly designated as *secondary hypotheses*. Note that the words "primary" and "secondary" are being used in a somewhat different sense than in the phrases "primary analysis" and "secondary analysis." An important reason for requiring investigators to specify their hypotheses in advance, and to only select a single primary research hypothesis, is to avoid the *multiple testing problem*, which is discussed in greater depth later in the chapter.

Not only must the main hypothesis be clearly stated in the study protocol, but also the outcome measure and statistical procedure used to assess this hypothesis must also be spelled out in advance (Friedman et al. 2010). For example, the researcher may specify that on the first day of each month of the follow-up period in a longitudinal study, each subject will self-report the number of drug use days over the previous month. Then, the effect of treatment on drug use level will be assessed by applying a random effects Poisson regression model to the outcome variable (monthly counts of active drug use days), thereafter performing a hypothesis test on the coefficient of the treatment status predictor variable. Finally, if the study design allows part of the data to be available for analysis prior to the collection of the remaining data, then any *interim analyses* that are conducted using only the initial available data must be specified in advance as part of the study design, and they must be accounted for in the statistical analysis plan.

5.2.2 Study Population and Sampling

A fundamental choice in the design of a study is selecting the study *population*, which is the set of all people we wish to make inferences about. This choice, however, cannot be solely based upon the population we are interested in learning about, as it may be difficult (or impossible) to randomly select an *unbiased sample* from this population.

For example, suppose a researcher wishes to choose all male heroin users over the age of 18 in Los Angeles County, California, as his/her population of interest, for which data will be collected to assess their research hypotheses. In order to take a *simple random sample* from this population (a sample for which all members of

the population have the same probability of being included in the sample), two things would be needed. First, would be an exhaustive list of *all* adult males in Los Angeles County who are current active heroin users. Second, for each of these subjects, the researcher would need a means of contacting the subject, *and* each and every subject must be willing to participate in our study. Of course, many heroin users are *hidden*, insofar as they may not have had contact with drug treatment facilities, law enforcement, or other public programs or organizations, which would allow them to be identified as heroin users. In addition, members of drug using populations often have unstable living situations, which make finding and contacting them difficult. Finally, subjects who refuse to participate in the study cannot be included in the sample. For these practical reasons, researchers must often restrict study populations to, for example, only subjects who have recently had contact with treatment facilities, and the percentage of contacted subjects who consent to participate should be reported so that the extent of possible selection bias may be assessed.

5.2.3 Sample Size Determination and Power Analysis

Sample size (N) is an important factor to all aspects of statistical work, including the stability of statistical estimation and the power of hypothesis testing. From a statistical viewpoint, the larger the sample size, the better the statistical inference. Unfortunately, the collection of sample data usually is time-consuming and expensive. Although possible (e.g., a census), the collected sample generally is much smaller than the population. Sample size determination usually is based on the balance between costs and statistical consideration (e.g., power and precision). In general, sample size should be appropriately determined before any prospective studies. In all retrospective studies, sample size is usually given and cannot be changed, so the power of a study needs to be reported to ensure the validity of statistical inference.

To determine the appropriate sample size for a specific study, there are several possible approaches. The most popular approach, however, is the power analysis for hypothesis testing approach. In this approach, a null hypothesis, H_0, and its alternative hypothesis, H_1 must be specified. Then, the sampling distributions of the predefined test statistic under both H_0 and H_1 must be derived, respectively. Given the sampling distributions and a predefined significance criterion, α, which is usually set to 0.05 for H_0, and the so-called effect size (d), determined by the null and alternative hypotheses, can be combined with sample size (N) to determine the power of the test statistic. The power of the test statistic here is defined as the conditional probability that the test statistic will *reject* the null hypothesis H_0 at the α level, given that the alternative hypothesis H_1 is true. In contrast, the conditional probability that the test statistic will *accept* H_0 at the α level, given that H_1 is true, is called type II error, which is usually denoted by β. Consequently, the power of the test statistic is denoted by $1 - \beta$.

In this power analysis for hypothesis testing framework, and given the other elements mentioned above, the effect size (d), the sample size (N) and the power ($1 - \beta$) typically can be determined if any two of them are specified. This relationship has some substantive relevance. In prospective studies, people are usually interested in determining the minimum N, which is necessary for the test statistic to detect an effect of a predefined size d with a desired amount of power for study planning ranging between 0.80 and 0.90. Sometimes the budget for the prospective studies can be limited, such that N must be restricted to a specific size. In this case, the focus usually changes to how large of the effect size can be detected in the study given the fixed N, and a desired power between 0.80 and 0.90. In retrospective studies, where N is fixed, people often focuses on the power of detecting the effect of a predefined size d. Of course, the focus can also be directed to how large of the effect size can be detected in retrospective study given the fixed N and a desired power, such as 0.80 and 0.90.

Using the one sample test of a single mean parameter μ as an example to demonstrate sample size determination and power analysis based on the framework above, imagine that a researcher is evaluating the effect of a new mathematic teaching method. They would need a random sample of N subjects from a population known to be normally distributed, with a μ of 100 and a σ of 15. Given the known population mean, the null hypothesis is $H_0{:}\mu_0 = 100$. For the study, the test statistic is the sample mean $\bar{X}$ of math test scores of N subjects after administration of the new teaching method. If it is supposed that subjects' true math ability (μ_1) after the new method has no real change, then $Z = \sqrt{N}(\bar{X} - 100)/15$ will follow a standard normal distribution (derived from the sampling distribution of $\bar{X}$ under H_0) and H_0 will be rejected at 0.05 of chance if $Z > 1.96$ or $Z < -1.96$, the critical values under H_0 when $\alpha = 0.05$. In this situation, $H_0 = H_1$, the effect size $d = (\mu_1 - \mu_0)/\sigma = 0$, and it is unnecessary to discuss the minimum N of detecting a zero effect size with a desired power such as 0.80 or 0.90.

Now if it is supposed that μ_1 has some real change, then $H_1{:}\mu_1 \neq 100$ and under H_1, Z will follow a normal distribution with a mean of $\sqrt{N} \times d$ and a standard deviation of 1 (derived from the sampling distribution of $\bar{X}$ under H_1). By this distribution of Z under H_1, the power or the chance of rejecting H_0 if $Z > 1.96$ or $Z < -1.96$ will increase when N increases and d departs from zero. Furthermore, μ_1 in H_1 could be any value other than $\mu_0 = 100$. It could be any value greater than 100 if the real effect is an improvement, and be any value less than 100 if it is a failure. For sample size determination and power analysis, μ_1 cannot be an arbitrary value and must be some specific value. This is generally achieved by defining the small/medium/large effect size for d based on Cohen (1988). Cohen (1988) defined $d = 0.20$, 0.50 and 0.80 as a small, medium, and large effect size, respectively. The rationale of this definition is that the small/medium/large effect size of d means the designated changes of ability $(\mu_1 - \mu_0)$ are the 20/50/80% of the common standard deviation, respectively.

The following definition makes the selection of μ_1 in H_1 less arbitrary in the given example. For example, now given d, μ_0, and σ, μ_1 by definition will be equal to 103, 107.5, and 112, respectively, corresponding to the small, medium, and large effect size defined by Cohen (1988). More importantly, this definition, in some degree, simplifies the procedure of sample size determination and power analysis. Any specific effect size d can correspond to an infinite combination of H_0 and H_1. For example, $d = 0.20$ could correspond to $(\mu_0, \mu_1, \sigma) = (100, 103, 15)$ in our example. It can also correspond to $(\mu_0, \mu_1, \sigma) = (80, 83, 15)$, or $(110, 113, 15)$. In the math teaching example, it is assumed that the population mean μ_0 of math ability are known so that μ_1 is determined by a defined d. Unfortunately, this assumption is often unrealistic in practice and the population parameter values, such as μ_0 and μ_1, are usually unknown in most cases. However, the effect size d defined by Cohen (1988) can represent an infinite combination of H_0 and H_1 and, at the same time, represents a substantively meaningful measure, the designated mean change $\mu_1 - \mu_0$ in terms of the proportion of the common standard deviation σ. Thus, by using the effect size d defined by Cohen (1988), sample size determination and power analysis not only are free from arbitrary choices of values for population parameters μ_0 or μ_1 but are also consistently related to a substantively meaningful quantity in practice.

In the example provided, given the sample distributions of $\bar{X}$ under H_0 and H_1, and the power set to 0.80 and $d = 0.20$, 0.50, and 0.80, then the minimum N needed is 197, 32, and 13, respectively. Note that this sample size determination still holds even when the population value of μ_0 or μ_1 is unknown. In other words, μ_0 and μ_1 could be any values but the mean difference $\mu_1 - \mu_0$ will be some fixed proportion of σ depending on the choice of d. For retrospective studies, with a fixed N and the given sample distributions of $\bar{X}$ under H_0 and H_1, the power can be determined for $d = 0.20$, 0.50, and 0.80; or in another way, the minimum detectable d can

be determined for a desired power, such as 0.80 or 0.90. Now suppose that the study included 50 subjects ($N = 50$). Then, the power to detect $d = 0.20$, 0.50, and 0.80 is 0.29, 0.94, and 1.00, respectively. Similarly, when the desired power is set to 0.80 or 0.90 and $N = 50$, then the minimum detectable d is 0.397 and 0.460, respectively. Again, note that this power analysis and effect size estimation holds even when the population value of μ_0 or μ_1 is not known.

The key concepts introduced here, such as the null and alternative hypotheses (H_0 and H_1), the sample distributions of the statistic under H_0 and H_1, and effect size d, apply to all kinds of power analysis and sample size determinations under the hypothesis testing approach. Among the applicable list, common examples include two-sample comparison of mean, simple regression/correlation, 1-way and factorial ANOVAs, general linear models, logistic regression, proportions, chi-square test of contingency tables, survival analysis, etc. For each case, the specific form of the key concepts mentioned above will be different. For the details of the differences and their implementation, Cohen (1988) is a good reference. At the same time, G*Power (Faul et al. 2007) is a free software that can be used with Cohen (1988) for power analysis and sample size determination.

Although very popular, the approach described above for sample size determination has its own limitations, for which the biggest challenge usually pertains to the sampling distribution of the test statistics. In some areas of statistics (e.g., finite mixture analysis, mediation effect analysis), the sampling distributions of statistics under H_0 and H_1 are barely derived. Without knowing the exact sampling distributions, the effect size, power, and sample size determination cannot be determined. For those areas, computational simulation could be a solution (Muthén and Muthén 2002). The values of population parameters, such as μ_0, μ_1, and σ, as well as N, are at first selected based on the empirical studies. Usually, several combinations of those parameter values (e.g., three levels of $\mu_1 - \mu_0$ and four levels of sample size N) will be used so that they can cover more experimental conditions that would be encountered in practice. Then the sampling distributions of statistics (e.g., $\bar{X}$) under H_0 and H_1 will be simulated computationally under various experimental conditions and the power under each condition will be discovered empirically. Finally, all results will be combined and provide some guidelines for real application. Of course, this approach has its own limitations. Simulation conditions are usually limited in scope and cannot cover all situations. This approach can be considered as remedy, rather than a replacement, to the hypothesis testing framework.

5.2.4 Understanding Study Characteristics for Secondary Analysis

Many aspects of study design have important effects on the kinds of questions the data collected from that study can be used to answer. For example, an experimental study in which study participants are randomly assigned to an intervention typically permits inference of a *causal* relationship between the intervention and the outcome variable, as opposed to an observational study, in which study participants decide what interventions they will participate in. In the following section, various aspects of how study participants are selected and how their data are recorded will be discussed, in addition to how these study aspects affect what secondary data analysts might use a study's data for.

5.2.4.1 Prospective Versus Retrospective Studies

The first aspect of study design we consider is whether the data we collect from subjects corresponds to measurements or traits that were present before or after the subject was recruited into the study. In a *prospective study*, data is only recorded at the time of recruitment or interview, or at a later time. On the other hand, a *retrospective study* uses records or interviews to obtain information about what happened prior to study initiation (Rothman et al. 2008). As such, each type of study has its own strengths and weaknesses. Since retrospective studies examine

previous outcomes, there is a danger that these outcomes will influence the chances of a subject being selected to participate in the study, which is commonly referred to as *sampling bias or selection bias*. Such a bias is less likely with a prospective study design, for if the outcome events have not yet happened at the time of recruitment, they cannot directly affect recruitment. In addition, data collected from interviews in retrospective studies may suffer from *recall bias*, which are inaccuracies in how subjects remember events in the past.

On the other hand, prospective studies (especially *longitudinal studies*, which are discussed below) can suffer from bias due to early participant dropout, called *loss to follow-up* or *attrition*, both of which relate to dropout occurring prior to the outcome being measured. For example, subjects who are increasing both the frequency and amount of their drug use may be less likely to report back to a facility for a follow-up interview. Another disadvantage of a prospective design, as compared to a retrospective design, is that if data needs to be recorded from the same subjects covering long spans of time, then the study must be conducted over long periods of time, whereas retrospective studies can immediately retrieve data from long spans of time in the past (though again, there will be an increased likelihood of *recall bias*).

5.2.4.2 Experimental Versus Observational Studies

As mentioned above, in *experimental studies*, the investigators *randomly assign* different *interventions* or *treatments* (such as different counseling programs or different pharmacological treatments) to study participants. The purpose of this treatment *randomization* is to ensure that groups of participants receiving different treatments are comparable, on average, with respect to other factors affecting the outcome variable being assessed (Friedman et al. 2010). To illustrate, suppose one wants to determine the effects of two different counseling programs, a single-visit program versus an intensive 8-week program. If, as in an *observational study,* the subjects choose which program to participate in, the more highly motivated subjects may choose the 8-week program, which would lead to exaggerated estimates of the *causal effect* of the intensive intervention. In this case, motivation would be a *confounding factor* for the relationship between intervention and the drug use outcome measure. In a secondary analysis of this data, the researcher would then need to *control* or *adjust* for this motivation factor in a multiple regression, as described later in Sects. 5.3 and 5.5. On the other hand, an experimental study would randomly assign the counseling programs to study subjects, thus making it unlikely that the intensive program subjects would be more motivated, on average, than the non-intensive subjects. In this case, no adjustment would need to be performed in the analysis. Finally, note that all experimental studies must be prospective (though they may retrospectively collect background information on subjects, such as age at first drug use), since investigators cannot have assigned treatment before the subjects were recruited.

5.2.4.3 Cross-Sectional Versus Longitudinal Studies

In many studies, it is of interest to determine how patterns of drug usage and related traits and behaviors change over time. In a *cross-sectional study*, a group of subjects is sampled and data is collected at a *single point in time*. Thus, the data from these subjects represent a *cross section,* or *snapshot* of people evolving through time. If interest centers on the relationship between drug use and a notion of time, such as age, time since first use of the drug, or time since the subject entered treatment, then inferences about these relationships may only be made *indirectly*, for example, by comparing people at age 30 to *different* people who are age 20. Differences between these people may be attributable to their age difference or to some other factor differing between these groups, such as ethnicity or primary drug type. On the other hand, in a *longitudinal study*, the *same* people are followed and measured at multiple occasions over time (Weiss 2005). In longitudinal studies, changes attributable to various notions of time may be made

directly, for example, by comparing people at age 30 to the *same* people when they were 20 years old. This prevents bias due to some potential confounding factor, such as ethnicity, but factors that may change through time, such as primary drug type, could still be confounders. Note that, technically, all longitudinal studies are prospective in design, though retrospective studies can collect *longitudinal data* corresponding to multiple time points in the past.

5.2.4.4 Drug Use and Other Outcome Measures

The type and number of outcome measures are often key determinants of deciding which statistical methods should be applied. The most important outcome measure in drug abuse research is drug use status, or level. Data for drug abuse research largely have been based on self-report, biological measures, or medical and administrative records. Self-reported drug use data could include type, frequency, or duration of use. The Addiction Severity Index (McLellan et al. 1992) is the most commonly used instrument, or measure, to indicate severity of drug use. Another common standard measure is diagnosis according to Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria. Self-report data are often criticized, however, as being subject to a variety of reporting biases due to unintentional (e.g., memory failure) and intentional (e.g., underreporting) factors. Biological measures, such as urine or hair testing results, are usually considered more reliable but have restricted time coverage. Medical and administrative records (e.g., arrest, death) are other objective data sources, but they do not usually provide drug use information. One challenge of secondary data analysis is to either select the best measure among several possible candidate measures that have been collected, or to derive a new measure from the variables available in the dataset. For instance, it may be reasonable to consider a subject as having a positive treatment outcome during the 12-month span following release from prison if self-report, urinalysis, and/or official medical and criminal records indicate drug abstinence or criminal desistance.

5.3 Regression Overview

In drug abuse research, interest often centers on the relationships between an outcome variable and predictor variables. Even if a researcher is only interested in the effect of a single predictor on the outcome, examining other variables may still be necessary to prevent bias in the effect estimate of interest. *Regression modeling* is the most common statistical method for assessing the relationships between predictors and a single outcome variable. However, it is important to note that the types of outcome variables determine the choice of regression type within (a) linear regression models applicable for continuous outcomes, (b) logistic regression models for categorical outcomes, and (c) survival analyses when outcomes of interest are a duration measure or time to the occurrence of an event.

5.3.1 Linear Regression for Continuous Outcomes

The type of regression model used will depend on the type of outcome variable being analyzed. A *linear regression* model is used when the outcome variable is measured on a continuous scale. For example, suppose the dataset to be used for a secondary analysis recorded, for each subject i, the amount y_i of alcohol consumed over the 1-year period following a treatment program, and this variable was selected as the outcome of interest. If predictor variables are $x_{1i} =$ age of subject i (in years at the time treatment begins) and $x_{2i} =$ treatment type (coded with $x_{2i} = 0$ if subject i participated in a single-visit program and $x_{2i} = 1$ for an intensive 8-week program), then the linear regression model would be:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \epsilon_i$$

where ϵ_i is an *error term* representing how much subject i differs from the average of subjects of the same age in the same treatment group. This would be a *multiple linear regression* model as it includes *multiple* predictor variables in the same

model; common issues that arise in multiple regression models are discussed in greater detail after other, nonlinear types of regression are explored.

5.3.2 Logistic and Poisson Regression for Categorical Outcomes

Suppose that instead of measuring a continuous value, such as the amount of alcohol consumed, the outcome measure y_i records *whether or not* subject i had some trait or performed some behavior. For example, $y_i = 1$ could be defined if subject i consumed *any* alcohol during the first month following treatment and, alternately, $y_i = 0$ if the subject was abstinent during that first month. In this case, a linear regression model for y_i would be inappropriate, since y_i can only take two values, 1 and 0, representing whether or not subject i consumed alcohol. Hence, the objective would now be to instead model the *probability* p_i that subject i consumed alcohol using a *logistic regression* model. This model says that the logarithm of the *odds* of alcohol use is a function of our predictor variables:

$$\log(p_i/(1-p_i)) = b_0 + b_1 x_{1i} + b_2 x_{2i}$$

Note that this model does not require an error term ϵ_i (Hosmer and Lemeshow 2000).

Alternatively, if the outcome variable y_i is a *count* of the number of days during the first month following treatment during which the subjects drank alcohol, then we would use a *Poisson regression* model. This model says that y_i has a Poisson probability distribution with mean μ_i, and this expected count μ_i is related to predictor variables via the regression relationship:

$$\log(\mu_i) = b_0 + b_1 x_{1i} + b_2 x_{2i}$$

As with the logistic regression model, this Poisson regression model does not include an error term (Agresti 2002).

5.3.3 Cox Proportional Hazards Regression for Survival Outcomes

Instead of measuring whether some event, such as alcohol consumption, occurs or the count of the number of occurrences of the event, one may be interested in *time until the event first occurs*. Such a variable y_i, recorded for each subject i, is often called an *event time* or *survival time* variable, since it measures the amount of time the subject "survives" free from experiencing the usually adverse event (Kalbfleisch and Prentice 2002; Klein and Moeschberger 2003). In this case, the variable y_i is continuous, so a linear regression model can be utilized. However, two important features distinguish survival time variables from other continuous outcome variables. First, the amount of time y_i that the subject remains free of the event cannot be a negative number, so the desire is for a regression model that does not allow $y_i < 0$.

More importantly, many subjects may never experience the event of interest during the course of the study, so it is unlikely to have a y_i value recorded for all participants. What is known is that if the event does ever occur for some such participant i, then $y_i > c_i$, where c_i is the amount of time to follow up with the subject to monitor him or her for event occurrence. For example, if a study for secondary analysis followed subjects for one year after they ended treatment, and participant i never relapsed into alcohol use over the course of the study, then all that is known about this subject is that his survival time $y_i > 1$ year. This common special feature of event time and survival outcomes is called *right censoring*, and it must be properly accounted for to obtain a correct analysis (Willett and Singer 1993).

The most common model for survival time outcome variables used in substance abuse research is the *Cox proportional hazards regression model* (Cox 1972). With this model,

the survival outcome, y_i, is modeled indirectly by modeling the *hazard rate* $\lambda_i(y)$, which represents chances that the event of interest occurs for the subject i at time y, given that the event has not already occurred previously:

$$\lambda_i(y) = \lambda_0(y)\exp(b_1x_{1i} + b_2x_{2i})$$

The function $\lambda_0(y)$ represents how the chances of event occurrence change given the amount of time (y) that the subject (i) has remained free of the event. The regression term $b_1x_{1i} + b_2x_{2i}$ (which we have written here with two predictor variables, though fewer or more variables are possible), allows us to model how the risk or hazard of event occurrence further depends on covariates, such as age, sex, and treatment type.

5.3.4 Multiple Regression and Extensions

A *multiple regression* is any of the above regression models that includes *multiple predictor variables*, such as x_{1i}, x_{2i}, x_{3i}, etc. Briefly discussed here are two common issues encountered when incorporating more than one predictor variable (or otherwise referred to as a *covariate*). These two issues are often conflated with one another.

5.3.5 Interaction, Moderation, and Effect Modification

The coefficients (b_1, b_2, b_3, etc.) in this multiple regression model tell us separately about the effects of changing one of the covariates (x_{1i}, x_{2i}, x_{3i}), while keeping all the others held constant. However, these coefficients do not allow the effect b_2 of a covariate x_{2i} to change, depending on the value that some other covariate x_{1i} takes. For example, suppose as before that $x_{1i} =$ age and $x_{2i} =$ treatment type. It may be believed that the effectiveness of the treatment intervention on curbing alcohol use depends on the age of the subjects, which would be an *interaction* between treatment and age (Rothman et al. 2008). One would then also say that age *moderates* the effect of treatment, and that age is an *effect moderator*.

A regression model with just x_{1i} and x_{2i} only allows a single effect estimate for each of these predictors, which does not depend on the value of the other predictor and so cannot capture interaction of the predictors. The solution is to create and include an additional predictor variable using the product of the two variables we believe may have an interaction, e.g., $x_{3i} = x_{1i} \times x_{2i}$. The coefficient b_3 then tells us how the effect of treatment x_{2i} changes with differing age x_{1i}, and so captures the interaction effect.

5.3.6 Confounding and Mediation

In secondary analyses, the analyst does not have control over the study design beyond being able to choose which dataset, or datasets, to include in the analysis. As mentioned in Sect. 5.2, if the dataset comes from an observational study, then one cannot necessarily infer that a variable has a *causal effect* on our outcome of interest merely from the fact that a regression model found a statistically significant *association* between that predictor variable and our outcome variable. It may instead be the case that a third variable, called a *confounding factor* (or simply a *confounder*), is responsible for the observed relationship between the predictor and outcome (Rothman et al. 2008). This is also possible in a secondary analysis of a randomized, experimental study, if the hypothesis of interest in the secondary analysis refers to a predictor variable that was not assigned using randomization in the experimental study.

An examination of the relationship between the number of days of marijuana use during a 12-month period following a treatment program and the type of treatment program that subjects have chosen to enroll in can be used as an illustration. Note that in this case, treatment is not randomly assigned by the investigators as in an experimental study, so factors other than chance alone may determine which treatment program

each subject participates in. If, for example, current legally employed subjects are more likely to choose an evening counseling program due to the inability to miss work during the day, then this program may appear more effective than the daytime program simply because employed subjects use marijuana less frequently than unemployed subjects. In this case, *employment status* is *confounding* the relationship between treatment program and the frequency of marijuana use over the 12-month study period.

Considering that in a secondary analysis the design of the study cannot be changed, one must attempt to *adjust for* or *control for* the presence of confounders after the fact in the statistical analysis. The most common way to do this is to include the confounder as an additional predictor variable in a multiple regression model. The coefficient of the original predictor variable of interest (treatment type in our example) will then represent the relationship between the predictor and outcome *given that the confounder is held fixed*, so that the confounder can no longer be responsible for the observed association between the predictor and response.

Because this method of accounting for potential confounding factors is so easy to implement (specifying additional predictor variables in multiple regressions in statistical software programs usually only requires typing in additional variable names), it is tempting to try to include any remotely plausible potential confounders "just to be safe." However, this can lead to two problems. First, including frivolous variables that are not associated with the response can decrease the precision of relationship estimates involving the predictor variables that are not of note. Second, accidental inclusion of a variable called a *mediator* can affect both the predictor of interest and the response variable.

For example, suppose as before that the number of days of marijuana use is the outcome of interest and the type of treatment program a predictor. Furthermore, suppose the dataset contains a survey response variable measuring the subjects' motivation to cease drug use. If the treatment program is a behavioral intervention designed to work, in part, by increasing subjects' motivation to reduce drug use, then *motivation level* would be *mediating* part of the causal effect of treatment. If one were to include the motivation-level survey response, in addition to treatment status in the multiple regression model, then the coefficient of the treatment variable would only represent the portion of the total causal effect of treatment that is not due to increasing motivation levels. Most likely, however, there is an interest in estimating the total causal effect that the intervention has on the response variable, including the portions mediated through other measured variables. In this case, motivation level should not be included as a predictor variable in the model, though it could be included in a mediational analysis, where the mediator's indirect effect can be estimated, in addition to the direct treatment effect.

5.3.7 Multilevel Modeling of Correlated Data

A basic assumption regarding a secondary dataset implicitly assumed in all the regression models discussed thus far is that any observations of the outcome within the sample are *independent* of one another given their covariate values. In other words, knowing the value of, say, y_i, tells us nothing about the value of y_j beyond the information contained in subject j's covariate values and the parameters of the model. This assumption can be violated, however, and the two most common types of secondary datasets that exhibit this correlation, or lack of independence, are as follows.

5.3.8 Repeated Measurements in Longitudinal Studies

As mentioned in Sect. 5.2, longitudinal drug abuse datasets often contain repeated measurements of the same subjects over time. For instance, a dataset may contain monthly summaries of drug use, such as the number of days of use that month, for each month of the first 5 years following treatment. Thus, for subjects who have

not dropped out of the study early, the dataset would contain 60 repeated measurements from each subject. If y_{ij} and y_{ik} are two such measurements, from subject number i, taken at months j and k, one would expect these measurements to be correlated with one another. In other words, if y_{ij} was a high value (e.g., $y_{ij} = 30$ days per month of drug use), then it is likely that $y_{i,j-1}$ and $y_{i,j+1}$ (the usage levels in the previous and following months, respectively) will also be high values.

The most common way to model correlation among repeated measurements in longitudinal datasets is to include an extra term, a_i, for each subject, i, in the regression model called a *random effect* (Weiss 2005). Then the regression model becomes, for example,

$$y_{ij} = b_0 + b_1 x_{1ij} + b_2 x_{2ij} + a_i + \epsilon_{ij}$$

in the case of a multiple linear regression with two predictors x_{1ij} and x_{2ij}. These predictors are then called *fixed effects* to distinguish them from the random effects, and the resulting regression model is called a *mixed effects* model, emphasizing the inclusion of both fixed and random effects. In addition, since conceptually the repeated measurements are *nested* within subjects in a hierarchy, the model is also sometimes called a *multilevel* or *hierarchical* model.

It is not necessary to find a value for each a_i among any of the observed values in the secondary dataset, since like the unknown parameters b_m, they will be estimated by the statistical model. The value a_i, which remains the same for all observations from subject i, allows the model to account for the similarity, or correlation, among the repeated measurements y_{ij} from subject i. If a_i is a high value, for example, then all of the measurements y_{ij} from that subject will tend to be higher than expected given their covariates.

5.3.9 Time Trend Modeling for Longitudinal Data

When conducting longitudinal data analysis, it is important to correctly model the relationship between the repeated measurements y_{ij} of the outcome variable and the time t_{ij} at which each measurement was taken. Thus, one may typically want to include t_{ij} as a predictor variable in the model. The model is then often called a *growth curve model*, because it is now modeling how the response variable y_{ij} changes, or "grows," over time. For example, suppose that the outcome variable increases *linearly* with time. Then, supposing two other predictor variables, x_{1ij} and x_{2ij}, as in the longitudinal mixed effects linear regression model above, could define the third predictor variable to be time, $x_{3ij} = t_{ij}$, in which case the growth curve model becomes:

$$y_{ij} = b_0 + b_1 x_{1ij} + b_2 x_{2ij} + b_3 t_{ij} + a_i + \epsilon_{ij}$$

Alternatively, if we think the relationship between time and the outcome is *curvilinear* (i.e., not a straight line), then we could also include *time-squared* by setting $x_{4ij} = t_{ij}^2 = t_{ij} \times t_{ij}$, giving the *quadratic* growth curve model:

$$y_{ij} = b_0 + b_1 x_{1ij} + b_2 x_{2ij} + b_3 t_{ij} + b_4 t_{ij}^2 + a_i + \epsilon_{ij}$$

Note that it is necessary to include the linear term t_{ij} when including the quadratic term t_{ij}^2 in the model.

Oftentimes the effect of a variable of interest, such as the treatment group, will not remain the same over the course of the study. Assume, for example, that x_{1ij} is subject i's treatment group, with $x_{1ij} = 1$ if subject i has been assigned to a new intervention and $x_{1ij} = 0$ if this person receives the standard treatment. Note that this means the covariate x_{1ij} has the same value for all repeated measurements j from subject i. The outcome variable y_{ij} is the number of days of drug use in month t_{ij}, where t_{ij} is the number of months since the assigned treatment was initiated.

If the first measurement y_{i1} is taken at time $t_{i1} = 0$ (that is, y_{i1} records the number of days of drug use in the month during which treatment was initiated), then one might expect this first measurement not to differ between the treatment groups on average because the interventions have

not had time to take effect. On the other hand, for measurements y_{ij} recorded at $t_{ij} = 6$ months since treatment initiation, one might expect the two treatment groups to have diverged on average, and by time $t_{ij} = 12$ they may have diverged even further with respect to our drug use outcome y_{ij}. If a term of the form $b_1 x_{1ij}$ is included in the model as in the regression equations above, assuming that b_1 is the average difference in the outcome between the two treatment groups, and, importantly, assuming that this average difference does not change over time. However, it is observed that the treatment group difference will likely start out small (perhaps at 0) and grow over time. Hence, instead of including the *main effect* term $b_1 x_{1ij}$ for the treatment group variable x_{1ij}, it may be desirable to include the *interaction* term $b_1 t_{ij} x_{1ij}$ between treatment group and time. Note that at time $t_{ij} = 0$ months, the interaction $t_{ij} x_{1ij}$ is zero, so the model will predict zero difference between the treatment groups, as desired. When $t_{ij} = 1$, the difference between the treatment groups will be b_1, when $t_{ij} = 2$, the treatment effect will be $2b_1$, and so on. Hence, the treatment effect here is modeled as affecting the *trajectories* of the drug use outcome over time, with the trajectories diverging further as time since treatment initiation progresses.

5.3.10 Clustered Subjects

Another common feature of substance abuse datasets is the clustering of subjects into groups. For example, the dataset in question may consist of subjects sampled from multiple treatment facilities. Subjects from the same treatment facility may be more similar to each other than subjects from different facilities, even after accounting for differences between facilities with respect to fixed effects, such as race, which we may be including in the regression model. This similarity might result from the fact that subjects at each facility are admitted according to the same admission criteria, or primarily come from a nearby geographical location.

As with the case of repeated measurements in longitudinal studies discussed above, the statistical model must account for the correlations among subjects sampled from the same cluster or group. This can be accomplished by using the same mixed effects, hierarchical modeling approach outlined for the longitudinal case, but with one important change. When using the model

$$y_{ij} = b_0 + b_1 x_{1ij} + b_2 x_{2ij} + a_i + \epsilon_{ij}$$

with clustered data, i now stands for cluster number (not subject number) and j stands for the jth subject within cluster i (not the jth repeated measurement within subject i).

This section comes to a close with an important word of caution regarding sample sizes in multilevel models. Assume the covariate x_{mij} does not vary within clusters; that is, it always takes the same value for all subjects within a given cluster. In such a case, the covariate does not depend on subject number j, but only the cluster number i, and we call $x_{mij} = x_{mi}$ a *cluster-level* covariate to distinguish it from a *subject-level* predictor, which may differ between subjects from the same cluster. In this case, the appropriate notion of sample size for purposes of estimating the effect of this covariate is the number of clusters, not the total number of subjects from all clusters.

To understand why this is the case, suppose a dataset consists of 100 subjects sampled from each of eight treatment centers, for a total of 800 subjects, and you wish to estimate the relative effectiveness of two treatment protocols, A and B. This may seem like a large enough sample, but assume now that five of the centers exclusively use protocol A and the remaining three only use protocol B. If subjects on protocol B appear to respond better to treatment, can it be concluded that protocol B is superior? Unfortunately, subjects on protocol B may respond better because, simply by chance, one (or two) of the protocol B centers was more effective in delivering care (aside from the protocol differences), or one of these centers had a more favorable patient population. In other words, the chances of one protocol appearing better simply because

that protocol had better centers is much more likely when the number of centers is small. If, on the other hand, there were 50 protocol A centers and 30 protocol B centers, then it is less likely that, by chance, B would appear superior due to the 30 protocol B centers having, for example, a more favorable patient mix on average. Thus, it is clear that for purposes of assessing the effect of a factor that does not vary within clusters, the correct notion of sample size is the number of clusters, not the total number of subjects. An analogous rule applies to longitudinal studies: for covariates that do not change over time within subjects (such as the subject's race), the correct notion of sample size is the number of subjects, not the total number of repeated measurements from all subjects.

5.4 Latent Variable Modeling

Latent variable models are widely used in secondary analysis within substance abuse research. Originally, latent variable modeling was a convenient way to represent hypothetical constructs (e.g., intelligence, motivation) that vary among subjects and are generally unobservable, but can only be measured by observed variables with error. With further advances, latent variable modeling became a broader way to represent unobservable (or indirectly observed) variation among statistical models. This could include random coefficients (e.g., random intercept or slope) in mixed-effect models, categorical latent classes in mixture models, latent factors specified as dimension reduction devices, or as hypothetical constructs in factor models or structural equation models. Given this property, latent variable modeling has become a growing area of application in substance abuse research.

5.4.1 Structural Equation Modeling

Structural equation modeling (SEM) is a multivariate statistical method used to estimate and test the relationship among multiple (observed, latent, or both) variables hypothesized by a series of linear or nonlinear (or both) regression equations simultaneously. In a traditional regression, such as $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \epsilon_i$, the relationship between the observed variables, such as y_i and x_{mi}, is specified (and also tested) by a single regression equation. Although used widely, it may not be adequate in some disciplines to discover the more complex relationships coexisting among more variables. For example, Hser et al. (2004) hypothesized a SEM model (see Fig. 5.1) on the relationships between drug treatment services, retention, and outcomes based on longitudinal data from the California Treatment Outcome Project.

In the model, the quantity and quality of drug treatment services at the 3-month follow-up,

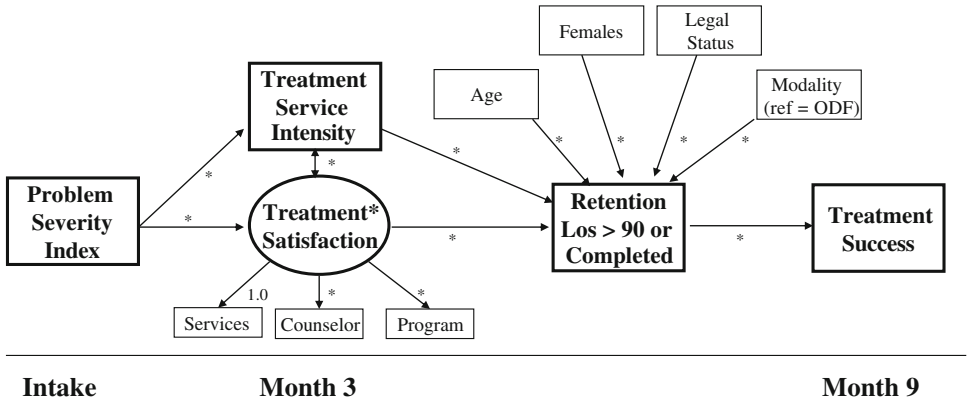

Fig. 5.1 Structural equation model in Hser et al. (2004)

represented in Fig. 5.1 by treatment service intensity and treatment satisfaction, respectively, were hypothesized to be influenced by the problem severity of subjects at intake and to be the predictors of the subjects' retention (i.e., treatment completion or continuing stay in treatment), along with other covariates such as gender, age, criminal justice status, and type of treatment program. At the same time, subject retention was also hypothesized to be the predictor of a favorable treatment outcome at the 9-month follow-up. In this SEM application, several regression equations must be specified and estimated simultaneously to evaluate the hypothesized relationships among the variables. In a typical SEM application, as in Hser et al. (2004), the modeling procedure often includes model specification, estimation, evaluation, and modification.

5.4.2 Model Specification

The specification of a series of regression equations for the variables under investigation is the first step of SEM modeling. In this step, substantive theory or previous empirical evidence are generally used to guide the model specification. Usually, the model specification is expressed in a diagram for ease of understanding and communication. Typically, as in Fig. 5.1, the variable in a square represents a single observed measure (e.g., problem severity index) and the variable in a circle represents a latent factor (e.g., treatment satisfaction) that is indicated by multiple observed measures (e.g., satisfaction with services, the counselor, and the program). The arrow indicates the direction of the hypothesized relationship, pointing from an independent variable (e.g., problem severity) to a dependent variable (e.g., treatment intensity) or from a latent factor (e.g., treatment satisfaction) to its indicators. A double arrow indicates a mutual influence or, technically speaking, a covariance between two variables (e.g., treatment intensity and satisfaction). The asterisks on the single and double-headed arrows represent the free parameters for estimation. In Fig. 5.1, the arrow from treatment satisfaction to one of its indicators is set to 1 to fix the scale of factor variance (the so-called *identification issue*; see Bollen 1989). The asterisk inside the circle denotes that the variance of the factor is free for estimation. In SEM, setting the variance of the factor rather than its path to one of its indicators as 1 is another way of fixing the scale of factor variance for identification purposes (see Bollen 1989).

In Fig. 5.1, the part of the SEM model without the indicators of treatment satisfaction specifies the path relationship between the variables and is often called the *structural part* of the SEM model, or structural model. If these factor indicators are omitted from the model and the treatment satisfaction factor is replaced by one of its indicators, the subsequent model is a path model. Similarly, the part of the SEM model that only includes treatment satisfaction and its indicators specifies the measurement of the latent factor and is often called the *measurement part* of the model, or measurement model. When this part is pulled out of the SEM model, the subsequent model is a single-factor model. In this sense, a regression model, path analysis, and factor analysis are all special cases of an SEM model.

To illustrate, suppose all regression relationships under study are linear. Let μ and Σ, respectively, be the mean vector and the covariance matrix of all observed dependent variables. In SEM, corresponding to μ and Σ, there is another mean vector and covariance matrix of all observed dependent variables, each of whose elements is implied by the hypothesized SEM model and can be expressed in terms of free parameters in Fig. 5.1. Let θ denote all those free parameters and $\mu(\theta)$ and $\Sigma(\theta)$ denote this model-implied mean vector and covariance matrix, respectively. Conventionally, for SEM model estimation and evaluation (see below), a null hypothesis based on the specified SEM model, as in Fig. 5.1, is often constructed as

$$H_0: \mu = \mu(\theta) \text{ and } \Sigma = \Sigma(\theta). \qquad (5.1)$$

The null hypothesis in (5.1) states that the specified relationship of observed dependent variables, as in Fig. 5.1, is the true relationship of

those variables. As a result, in terms of their means and covariances, the structured mean vector $\mu(\theta)$ and covariance matrix $\Sigma(\theta)$ should be equal to the unstructured μ and Σ that hold in the population and are independent of any model. Based on this null hypothesis of the structure of μ and Σ, further estimation and evaluation of SEM models can be conducted. Due to this reason, SEM modeling is often called *mean and covariance structural analysis*.

The mean structure $\mu(\theta)$ theoretically irrelevant at times, such as the case of the path analysis and factor model, and is often omitted from SEM model specification. As such, the null hypothesis then becomes $H_0{:}\Sigma = \Sigma(\theta)$. This type of SEM modeling is often called *covariance structural analysis*. Of course, some SEM models need the specification of $\mu(\theta)$ absolutely. For example, the latent growth curve model (e.g., Bollen and Curran 2006), which is another special case of the SEM model, assumes some growth factors (e.g., intercept and slope factor) to model the trajectories of repeated responses and needs its specification.

In SEM, a specified model sometimes cannot be estimated due to the problem of identification, which must be addressed during the model specification stage. A model is identified if we are able to obtain a unique solution for every free parameter. In Fig. 5.1, one of the paths from treatment satisfaction to its indicators is fixed as 1, to set the scale of latent factor variance. When $\mu(\theta)$ is included, the scale of the latent factor mean needs to be set for identification too. Usually, one of the intercepts of those factor indicators is set to 0 for this purpose. In addition to the scale of factor mean and variance, there is another necessary condition for identification. Let p denote the number of dependent variables. The degrees of freedom of the specified SEM model is equal to the total number of unique elements in μ and Σ, which equal $p + p(p+1)/2$ minus the number of free parameters in θ. As the necessary condition for identification, the degrees of freedom of the SEM model must be greater than or equal to zero. Depending on the model complexity, more rules may be required at times (Bollen 1989). In practice, once the rules listed above are fulfilled, the easiest way to determine identification is to run the specified model in some common software (see below) and check the error messages.

5.4.3 Model Estimation

In practice, the unknown parameters θ have to be estimated from the data. Under the null hypothesis in (5.1), many estimation methods were developed. Some common methods include the maximum likelihood (ML) method, generalized least-square (GLS) method, and asymptotically distribution-free (ADF) method. These estimation methods differ in their assumptions about the data. Both the ML and GLS methods assume multivariate normality of the observed variables. In contrast, the ADF method has no such requirement. Browne (1974, 1984) demonstrated that all three methods could be considered as a weighted least-square procedure with different weights that are based on their own assumptions of the data. Due to its specific weight, the ADF method often needs a very large sample size (at least 1000) and could be impractical for substantive research. The GLS method uses the sample mean and covariance matrix as its weight, and its stability could be influenced greatly by the characteristics of the sample (e.g., small sample size, excess kurtosis). The ML estimation is by far the most popular and recommended method in the literature. It often gives the most efficient (minimum variance) estimate. In many cases, it is less vulnerable to the characteristics of the data and violation of the multivariate normality assumption.

Satorra and Bentler (1988, 1994) provided a robust procedure, based on the weighted least-square framework, for the standard errors and the chi-square test statistics (see below) from ML and GLS estimation. The adjusted standard errors more precisely reflect the uncertainty of the ML and GLS estimators under the violation of normality assumptions. On the other hand, by using the ML estimator, its favorable properties (e.g., the robustness to small sample size) are retained. The ML estimation, coupled with the

Satorra–Bentler adjustment procedure, ultimately works better than many other methods across more realistic data conditions (e.g., small sample size, missing, and non-normal data) and has become a standard procedure of SEM estimation throughout the literature base.

All of these SEM estimation methods need some iterative process to obtain their estimate of θ. Common commercial software packages for SEM estimation include LISREL 8.8 (Jöreskog and Sörbom 2006), Eqs. 6 (Bentler 2000–2008), Mplus 7 (Muthén and Muthén 1998–2012), SAS PROC CALIS (SAS Institute Inc. 2013), and SPSS AMOS 7.0 (Arbuckle 2006). Free software packages include R (e.g., lavaan package) and Mx 6.0 (Neale et al. 2003). Most of these software packages (e.g., LISREL, EQS, Mplus, R lavaan) include the Satorra–Bentler correction procedure for the ML estimation.

5.4.4 Model Evaluation

SEM model evaluation, like its estimation, is based on the null hypothesis in (5.1). The evaluation of the hypothesized model, as in Fig. 5.1, is achieved by testing the tenability of the null hypothesis in (5.1). Many test statistics were developed for this evaluation. Usually, those test statistics have their corresponding estimation methods and follow a chi-square distribution. For example, the ML chi-square test (also known as the "maximum likelihood ratio test") is based on the ML estimation, and the ADF chi-square test is based on the ADF estimation. Like their estimation counterparts, these test statistics can perform well or poorly, depending on the situation. The ML and GLS chi-square tests require an assumption of multivariate normality and are vulnerable to its violation. Although it does not require an assumption of normality, the ADF chi-square test, like the ADF estimation, requires a very large sample size (>2500). The chi-square tests, based on ML and GLS estimation, can also be adjusted by the Satorra–Bentler procedure to make them less vulnerable to violation of the normality assumption and retain the favorable properties of their estimators at the same time. The so-called *Satorra–Bentler scaled test* is the mean adjusted ML chi-square test. It has been justified by many studies under various situations and was considered a standard test for the null hypothesis in (5.1). Most software packages (e.g., LISREL, EQS, Mplus, R lavaan) provide the Satorra–Bentler scaled test in their output.

In SEM, testing the null hypothesis in (5.1) is just one approach of model evaluation. This approach to evaluation assumes that the specified model, as in Fig. 5.1, holds exactly in reality, and that the corresponding test statistics (e.g., the Satorra–Bentler scaled test) evaluate this exact fit. However, this exact fit evaluation is unrealistic in practice. The chi-square tests are often very powerful and almost exclusively reject any model, even with a tiny misspecification, as the sample size increases. As a result, another approach to model evaluation, the so-called *close-fit evaluation*, was proposed in SEM. Instead of assuming the null hypothesis in (5.1), this approach evaluates how closely the specified model approximates the truth. More specifically, the target of evaluation is not the equality between the unstructured μ and Σ and the structured $\mu(\theta)$ and $\Sigma(\theta)$. Instead, the degree of approximation of the structured $\mu(\theta)$ and $\Sigma(\theta)$ to the unstructured μ and Σ becomes the target.

Numerous fit indices were developed for the close-fit evaluation. Among a long list of those fit indices, the most common include: the Bentler's Comparative Fit Index, or CFI (Bentler 1990), the Tucker–Lewis index, or TLI (Tucker and Lewis 1973), the root mean square error of approximation, (RMSEA) (Steiger and Lind 1980), and the standardized root mean square residual, (SRMR) (Bentler 1995). In SEM, none of these popular fit indices alone is adequate for the close-fit evaluation of a model. Instead, they should be examined and reported jointly for such a purpose. For example, the following criteria are a good rule of thumb for a good fit: TLI > 0.95, CFI > 0.95, RMSEA < 0.06, and SRMR < 0.08 (Hu and Bentler 1999; Kline 1998).

5.4.5 Model Modification

The fit of the original specified SEM model is often inadequate in practice, even by the close-fit evaluation. A strategy to deal with this inadequacy is further modification of the model (e.g., adding more parameters) until its fit indices become acceptable. The most common statistical method for this purpose is modification indices, which is also called the *Lagrange Multiplier Test* in EQS (Chou and Bentler 1990; Sörbom 1989). Most software provides modification indices to assist the modification. Modification indices indicate the expected improvement in overall model fit after one or more parameters (e.g., paths or covariances free for estimation) are added into the inadequate model. This information often drives the modification process sequentially until the modified model reaches some satisfactory fit. This exploratory strategy could be dangerous, however, for without cross-validation, the modified models obtained could be data-driven ones that are not generalizable across samples (Chou and Bentler 1990; Green et al. 1998; MacCallum et al. 1992). Simulation studies (Homburg and Dobartz 1992; MacCallum 1986) indicate that these specification searches can fail to uncover the correct underlying model, particularly, when the original model has many specification errors, when the sample size is small, and when the search is guided solely by a desire to improve the overall fit of the model.

Despite these criticisms, model modification without any cross-validation is still used in many SEM applications. One major reason for this is that data can be expensive to obtain and discarding it completely without any "discovery," or gain in knowledge, is not the wisest thing to do. An alternative modeling strategy for model modification is to use multiple a priori models. One or more theoretically plausible models representing competing hypotheses are specified and evaluated, in addition to the original model. This approach has been advocated by many (e.g., MacCallum and Austin 2000; Weston and Gore 2006) and faces less criticism.

5.4.6 Latent Class Analysis

Latent class analysis (LCA) is a statistical tool to identify distinct, unobserved groups (latent classes) of related data from multivariate categorical data. For example, in a recent study on cocaine-dependent men, LCA was chosen to investigate their heterogeneity on self-efficacy (Dang 2011). Self-efficacy across eight different types of situation categories, including (1) unpleasant emotions, (2) physical discomfort, (3) pleasant emotions, (4) testing personal control, (5) urges and temptations to use, (6) conflict with others, (7) social pressure to use, and (8) pleasant times with others was evaluated at intake, as well as 1 year and 2 years following treatment. Instead of assuming that the participants belonged to a single group, LCA was used to determine whether the cocaine-dependent men were relatively homogeneous in terms of their self-efficacy across different situations, or whether there was adequate unobserved population heterogeneity such that participants could be classified into distinct groups based on their patterns of responses on the eight context-specific domains of self-efficacy.

5.4.7 Model Assumptions

To achieve its purpose, latent class analysis assumes that the observed categorical variables are independent of each other within each unobserved group (local independence assumption), and that the probability of the observed values of each variable could vary across groups. Latent class models with different numbers of classes are compared to determine the number of classes that adequately capture the heterogeneity among the patterns of responses. With the assumption of latent classes as a tool for classification, LCA is considered to be a special case of finite mixture analysis. On the other hand, the local independence in LCA is similar to the independence of observed variables, given the latent factors in factor analysis. This similarity suggests that LCA is an analog of factor analysis

for categorical data. Specifically, with local independence, the within-class homogeneity and the across-class heterogeneity coexist, thus suggesting that the association between observed categorical variables in LCA is completely explained by the latent variables. From this perspective, LCA, like factor analysis, is a measurement model with latent categorical factors (latent classes) measured by their categorical indicators (observed variables).

In LCA, all observed variables are categorical so that LCA makes no special assumptions about the distributions of the observed variables other than that of local independence. The specification of LCA includes identifying two sets of parameters: class membership probabilities and item-response probabilities conditional on class membership. Multinomial distribution is often used to model both the class membership probabilities and the conditional item-response probabilities. Like factor analysis, the LCA model is often drawn as a diagram, and is very similar to the one on factor analysis. Typically, each latent class, denoted by a circle, has arrows pointing to each of its indicators in the square and the double-headed arrows among each latent class pair denote their association.

5.4.8 Model Estimation

The parameters of LCA are usually estimated by the maximum likelihood (ML) method. The expectation–maximization (EM) and Newton–Raphson (NR) algorithms are two of the most popular ML estimation methods. The EM method is a stable iterative method for ML estimation with incomplete data. The NR method is a faster procedure that directly uses the matrix of second-order derivatives of the log-likelihood function, which is needed to obtain standard errors of the model parameters. Both methods (especially the NR method) need good starting values to converge to the global maxima of the log-likelihood function. In practice, one way to proceed is to estimate the model with different sets of random starting values. Typically, several sets converge to the same highest log-likelihood value, which can then be assumed to be the ML solution.

The common commercial software packages for LCA estimation include Latent GOLD 5.0 (Vermunt and Magidson 2013) and Mplus 7 (Muthén and Muthén 1998–2012). Free LCA packages include LEM 1.0 (Vermunt 1997), MLLSA 4.0 (Eliason 1997), R (e.g., MCLUST and poLCA package), SAS Proc LCA 1.3.0 (Lanza et al. 2013), and WINMIRA (von Davier 1997). Most software packages automatically generate different sets of random starting values to ensure the adequacy of the estimation.

In LCA estimation, model parameters sometimes may not be identifiable. Different sets of parameter estimates can yield the same maximum of the log-likelihood function so that there is no unique solution to every free parameter. There are no general rules for the identification of LCA models. A necessary condition for identification is that the number of degrees of freedom must be greater than or equal to zero. But it is not a guarantee. For example, one needs at least three indicators for the LCA model, but if these are dichotomous, no more than two latent classes can be identified. With four dichotomous variables, the three-class model is not identified, even though it has a positive number of degrees of freedom. With five dichotomous indicators, however, even a five-class model is identified. For substantive researchers, the easiest way to determine identification again is to run the specified LCA model in some common software with different sets of random starting values and check the error messages.

In LCA, it is possible to achieve identification by constraining certain model parameters. For example, constraining the item-response probabilities to be equal across indicators can reduce the number of free parameters and make a two-class model with two dichotomous indicators identifiable. Imposing constraints can also prevent the occurrence of some other estimation problems (e.g., boundary solutions). In practice, prior information on model parameters from previous studies can be used to impose those constraints for estimation purpose.

5.4.9 LCA Class Enumeration

For LCA models, the correct number of classes must be selected to adequately capture the heterogeneity among the patterns of responses. This class enumeration is a key issue in LCA, as well as other mixture models described below. After all, distinction of groups represents a qualitative difference among subjects. Without accurate information on the distinctive groups, further discussion of quantitative difference would be less meaningful.

Many statistical indexes are available for the LCA class enumeration. These indexes fall into three general categories: (a) information criteria (IC) statistics, (b) entropy-based indexes, and (c) likelihood ratio test derivatives. Popular information criteria include Akaike's Information Criterion, or AIC (Akaike 1987), the Bayesian information criterion, or BIC (Schwartz 1978), and the sample size adjusted BIC, or saBIC (Sclove 1987). The common entropy-based indexes include the normalized entropy criterion, or NEC (Celeux and Soromenho 1996), the classification likelihood criterion, or CLC (Biernacki and Govaert 1997), and the integrated-completed likelihood criterion, or ICL-BIC (Biernacki et al. 2000). Likelihood ratio test derivatives include the bootstrapped likelihood ratio test, or BLRT (McLachlan 1987) and the Lo–Mendell–Rubin likelihood ratio test, or LMR LRT (Lo et al. 2001). For class enumeration, the specified LCA model with different number of classes is estimated first. Then, for each of the IC statistics, the lowest value among this series of models indicates the optimal number of classes. For entropy-based indexes, the same decision rule is used. For LRTs, the p-value of a k class model is used to test the fit of a $k-1$ class model versus k class LCA model, and a low p-value (e.g., less than 0.05) indicates that the $k-1$ class model has to be rejected in favor of a model with at least k classes. For LCA class enumeration, several studies (Nylund et al. 2007, Yang 2006) demonstrated that SABIC and BLRT are more reliable than the other indexes.

In addition to the statistical indexes above, substantive checking of the estimated LCA model is also an important source for class enumeration (Muthén 2003). After all, a model without meaningful substantive interpretation is useless. In LCA, model convergence and improper solutions can be avoided by adding constraints. However, they are also the indicators of overextraction of latent classes. Sometimes, the extracted classes with an extremely small size or proportion are also an indication of overextraction. In sum, class enumeration needs to combine different sources of information and can sometimes be considered an art in practice.

5.4.10 Covariates and Distal Outcomes

With a justified number of classes on the patterns of responses, subjects are often classified into more homogeneous groups based on their estimated posterior probabilities in many LCA applications. Thereafter, background characteristics of subjects and some other outcome variables at either follow-up or later stages (distal outcomes) are compared among estimated groups via chi-square, ANOVA, or multinomial logistic regression to estimate the covariate effects on the patterns of responses and the group differences on distal outcomes. This approach to the investigation of covariates and distal outcomes is often called the *classify-analyze approach*. In LCA, an alternative method is the one-step approach, in which the LCA model is specified to include covariate effects on latent classes and prediction of distal outcomes by latent classes as the parameters in a single overall model. Then, the specified overall model is fitted to the data to determine the number of classes, while simultaneously investigating covariate effects and group differences on distal outcomes (Muthén and Muthén 1998–1998).

5.4.11 Latent Transition Analysis

Latent transition analysis (LTA) is an extension of latent class analysis over time. In our self-efficacy example, latent class analysis is applied to

self-efficacy across eight different types of situations at each of three time points: at intake, 1 and 2 years following treatment. As a result, subjects with similar self-efficacy are grouped at each time point in the LCA model. LTA, as an extension, is further used to understand how these group classifications change over time, as well as to what extent other variables can impact this change. Instead of the LCA's classify-analyze approach, LTA investigates group classifications and their changes over time in a single model. In this model, groups of self-efficacy measured by the eight context-specific domains of self-efficacy at each time point (LCA models) are treated as discrete states and are estimated simultaneously with the transitions between these discrete states over time points (Collins 2006; Vermunt 2004). As illustrated in Fig. 5.2, at intake, 1 and 2 years following treatment, there is an LCA model in which groups of self-efficacy (denoted by Latent Class *i*) are measured by context-specific domains of self-efficacy. The arrows between the groups of self-efficacy represent the transitions between states over time.

As an extension, LTA has some assumptions in addition to the ones in LCA (e.g., local independence). LTA specifies three sets of parameters: item-response probabilities conditional on class membership at each time point, class membership probabilities at intake, and transition probabilities between states since intake. One key assumption is that the transition probabilities between states from time t to $t+1$ only depend on the available information at time t, the prior earlier time, and not on the ones at any additional times. Given this assumption, LTA clearly is a type of first-order Markov model, and the transition probabilities in practice are often modeled by some logistic models conditional on class membership and covariates at time t. Moreover, in addition to local independence at each time point as in LCA, LTA assumes that the indicators across different time points are related to each other only through the latent classes. Although not required, the number of states over time points usually is assumed to be the same in most LCA applications.

The estimation of LTA usually is completed via the EM algorithm. The common software packages include Latent GOLD, Mplus, LEM, R (e.g., CATLVM package), and SAS Proc LTA 1.3.0 (Lanza et al. 2013). Most software packages automatically generate different sets of random starting values to ensure the estimation. LTA, like LCA, also has identification issues in estimation, and there are no general rules for the identification of LTA models. Again, a necessary condition for identification is that the number of degrees of freedom must be greater than or equal to zero, but identification is not a guarantee. For substantive researchers, the easiest way to determine identification, as per above, is to run

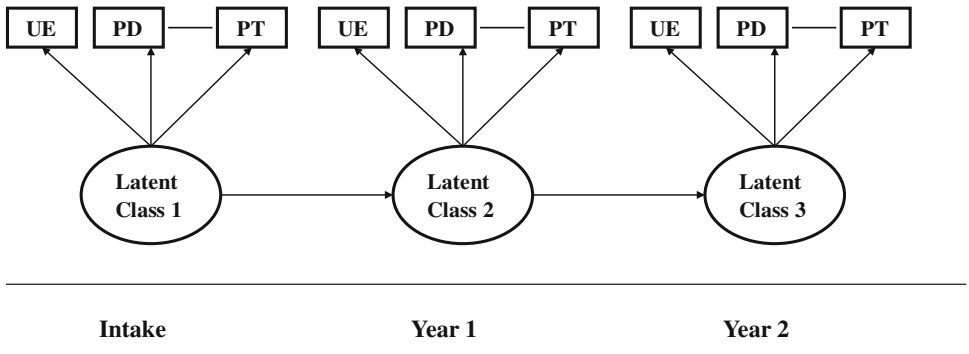

Fig. 5.2 Latent transition model (Dang 2011)

the specified LTA model in some common software with different sets of random starting values and check the error messages.

In many LTA applications, the measurement of LCA models at each time point often are first evaluated by various statistical indexes to determine the number of classes. Then with the appropriate number of classes at each time point, the overall LTA model is further estimated. In some applications, the specified LTA models with different number of classes are directly fitted to the data. In LTA, competing models are often evaluated and compared based on the likelihood ratio statistic, AIC, BIC, and interpretability of the latent classes.

When more than two time points are under investigation, the same probability model (usually a logistic model) is often assumed for the transition probabilities in the sequential stages. This constraint can be released, and then the LTA model becomes more complicated for estimation. In LTA application, some transition probabilities may be constrained to zero when those transitions between states are very unlikely. Constraining those parameters, as well as other parameters of LTA models can simplify the model and help the identification of the model. In LTA, the constraints (either freeing or fixing parameters) are usually evaluated by the likelihood ratio statistic. Like LCA, covariates' effects on class membership at intake, transition probabilities after intake, and prediction of distal outcomes by latent classes can be incorporated into the LTA model to make the model more interpretable.

5.4.12 Growth Mixture Modeling

Growth mixture models, or GMM (Muthén and Muthén 2000; Muthén and Shedden 1999; Nagin 1999) are a statistical tool to discover unobserved heterogeneity of growth trajectories among subjects and identify distinct groups (latent classes) underlying heterogeneity. In a 33-year follow-up study of 471 heroin users (Hser et al. 2007), GMM analysis was applied to identify groups with distinctive heroin use trajectory patterns during the first 16 years of their heroin addiction careers. In modeling the continuous outcomes of heroin use over time, distinct groups of trajectories were hypothesized and random intercepts, random slopes, and random quadratic growth terms were further specified within each group to fully capture the heterogeneity within the hypothesized groups. The study finally identified three trajectory groups (see Fig. 5.3 for the average observed heroin use levels for participants in the three trajectory groups). The group with the largest number of participants was the stably high-level users (n = 278, 59%), who maintained a fairly consistent high level of heroin use since initiation. The second largest group was the late decelerated group (n = 149, 32%), who maintained a high level of use for approximately 10 years, but then the percentage of nonusers started to increase, whereas those not quitting remained at their high level of use. The early quitters group had the smallest number of members (n = 44, 9%). These participants decreased their use within 3 years of initial use and stopped using altogether in the subsequent 7 years. In this study, the

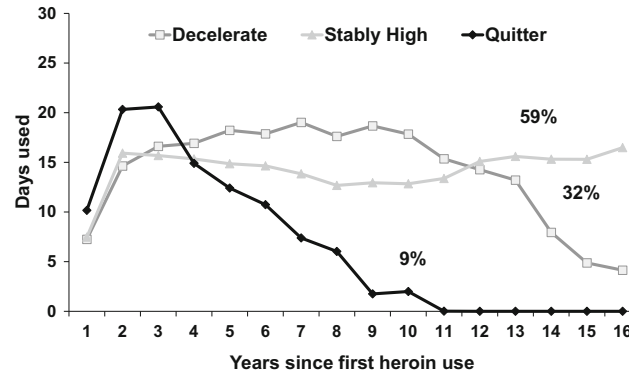

Fig. 5.3 The average observed heroin use levels for participants in the three trajectory groups

classify-analyze approach was used to further investigate group differences in the characteristics of participants (e.g., demographics, early deviance and problems in family and school, age of initiation of substance use, and arrest) and their outcomes at the 33-year follow-up (e.g., mortality).

5.4.13 Variations of Growth Mixture Models

There are many variations of growth mixture models. In one branch, trajectories are considered to be homogeneous within groups so that no additional random effects (e.g., random intercepts and slopes) are needed for trajectories within each group. This school of GMM modeling is called *group-based trajectory analysis* (Jones et al. 2001; Nagin 1999; Nagin and Tremblay 2001). For this school of models, several parametric models (e.g., censored normal model, zero-inflated Poisson model) were developed for various distributions of outcomes over time (Jones et al. 2001). Proc TRAJ (Jones et al. 2001) is the common software for this kind of modeling.

Another branch of GMM models (Muthén 2004; Muthén and Muthén 2000; Muthén and Shedden 1999) assume both latent classes for distinctive trajectory groups and random effects (usually growth factors such as intercepts and slopes factors) for the within-class trajectories. For this school of models, the parametric models for various distribution of outcomes (e.g., normal-component GMM model, censored normal model, zero-inflated Poisson model) have also been developed (Muthén and Muthén 1998–2012). With the existence of growth factors in those models, the variances/covariances of growth factors and residuals could be specified as invariant or noninvariant across different classes. Consequently, there are both class-invariant and class-specific GMM models. For this school of models, Mplus 7.0 (Muthén and Muthén 1998–2012) is currently the most common software for modeling.

Unlike LCA and LTA, GMM models usually can be identified. The more serious issue for GMM models is the quality of convergence. The solution obtained through an iterative process may be some local maximum and may not be the true ML solution. This is especially true for class-specific GMM models. As a result, a different set of starting values (e.g., in Mplus) is usually needed in GMM estimation to ensure the quality of convergence. Like LCA and LTA, constraints (e.g., constraining the variance/covariance of growth factor to be more invariant across classes) can help convergence, although these constraints usually need substantive justification and statistical evaluation by the likelihood ratio test.

5.4.14 GMM Class Enumeration

Nagin (1999) and Nagin and Tremblay (2001) recommended BIC for class enumeration of the group-based trajectory models. For the school of models including within-class heterogeneity, the three groups of statistical indexes available to the LCA class enumeration can also be used for GMM class enumeration. Nylund et al. (2007) suggested the use of BIC and BLRT for class enumeration. Tofighi and Enders (2007) found that saBIC is more accurate when the true GMM model has completely different variances/covariances of growth factors and residuals. Recently, both Li and Hser (2011) and Peugh and Fan (2012) found that the IC statistics and likelihood ratio statistics could diverge on the number of classes in more complex situations for class-invariant GMM models, and the true number of classes usually lies between the numbers of classes suggested by two branches of statistical indexes. This disparity again reflects the complexity of GMM class enumeration and can provide some clues for the true number of classes in more realistic situations. As in LCA, substantive checking of the GMM models is another important source for class enumeration (Muthén 2003). Model convergence and improper solutions are usually good indicators of latent class over-extraction (Tofighi and Enders 2007). At times, extracted classes with an extremely small size or proportion are also indication of overextraction.

5.4.15 Covariates and Distal Outcomes

The classify-analyze approach, as was used in Hser et al. (2007), is widely used for the investigation of covariate effects and prediction of distal outcomes in GMM applications. Sometimes, covariates and distal outcomes are incorporated into the GMM model as a single model for investigation. In such a model, covariates usually predict class membership, while growth factors within classes (if assumed) and distal outcomes are usually predicted by the latent classes and some covariates. Sometimes, time-varying covariates are also incorporated to predict the outcomes directly within each class. Although this overall GMM model with inclusion of covariates and distal outcomes can be used to determine the number of classes, it is often treated as a modification to the model prior to the inclusion.

5.5 Special Issues in Secondary Analysis

Several unique statistical issues arise when reusing data in secondary analysis. The first concerns the consideration of different hypotheses in the new data analysis. The second commonly encountered issue is how to combine results from multiple studies or datasets to address a single hypothesis and form a single, overall conclusion. The third problem concerns missing data and factors, and ways to handle them. The fourth concerns pretreatment differences or selection bias and the use of propensity-score methods for bias adjustment.

5.5.1 Multiplicity of Secondary Hypotheses

As mentioned in the introduction, usually a secondary data analysis is conducted to assess a different hypothesis from the one the original study was designed to assess. In many cases with large datasets, such as randomized controlled clinical trials, the original study designers are required to clearly specify their primary scientific hypothesis of interest in advance of data collection, and so they are not at liberty to select or modify their hypotheses after obtaining the data. With secondary analysis, however, there is often nothing to prevent a researcher from changing his or her hypotheses based upon what appears in the dataset. A danger then arises that the dataset will be used to assess numerous hypotheses, and only those with "positive" results (large, statistically significant effects) will be reported.

Unfortunately, it is well known that checking many hypotheses increases the chances that at least one of the significant relationships is a spurious finding. There are two main ways to address this problem. First, one can clearly pre-specify a main research hypothesis and, in addition, specify the variables from the dataset and the statistical model that will be used to carry out the test associated with the hypothesis. Any other research questions to be addressed should then be clearly specified as ancillary hypotheses (these are also often called *secondary hypotheses* in contexts where it is not necessary to distinguish primary and secondary data analyses). All results, positive and negative, of any of the pre-chosen hypotheses should then be included in the reported findings of the secondary data analysis. The second way to address the problem of spurious findings when considering numerous hypotheses is to use special statistical adjustments.

5.5.2 Statistical Adjustment for Multiple Comparisons

If a researcher conducting a secondary analysis wishes to assess several main hypotheses, then it is possible to use statistical adjustments, called *multiple comparison procedures,* to prevent inflation of the chances of a false positive finding. In general, the researcher first must select the

desired overall statistical significance threshold α (often called the *type I error rate*) and then carry out modified versions of the corresponding statistical tests for each of the pre-specified main hypotheses. The most commonly used of these procedures is the *Bonferroni correction*, which may be applied to any prespecified finite collection of hypotheses tests, even if different statistical models or tests are used for the different hypotheses. If M denotes the number of hypotheses, then given the prespecified overall type I error rate α (usually $\alpha = 0.05$ or $\alpha = 0.01$), we conduct each of the M tests separately using an adjusted significance threshold of $\alpha^* = \alpha/M$.

To illustrate, suppose one is conducting a secondary analysis of a study that followed subjects arrested for drug offenses after release from jail or prison. Of interest is the relationship between age at the time of release and ability to reintegrate into the community. Also of interest are the events of relapse into drug use and rearrest. It is possible to fit two Cox proportional hazards models using *time to relapse* and *time to arrest* as the survival time outcome variables, with each including the variable age among the predictors. Using $\alpha = 0.05$, each of the coefficients of the age variables from the two Cox models at the $\alpha^* = 0.05/2 = 0.025$ level are tested (since we are conducting $M = 2$ tests). Finally, age would be declared to be significantly related to either relapse or recidivism only if the corresponding test was significant at the 0.025 level.

5.5.3 Using Data from Multiple Studies

Thus far, we have discussed the reanalysis of data from a single study. However, it may be desirable to reuse data or results from multiple studies at once. This can allow us to have a larger effective sample size or a sample representative of a broader population than is possible when reanalyzing a single study's data. In this section, we discuss the most common approaches for incorporating data from multiple studies to reach a single conclusion regarding a single hypothesis.

5.5.3.1 Systematic Reviews and Meta-analysis

The simplest and easiest way in which the results of multiple studies may be reused is to conduct a literature review of published analyses relevant to the current hypothesis of interest. However, classical literature reviews suffer from three problems. First, the conclusions drawn from the review may be biased by the selection of papers included in the review, and second, these conclusions consist of *qualitative* summaries of the data analyses included in the review, not *quantitative* results, such as overall effect estimates, confidence intervals, or p-values. Finally, the scientific questions that can be answered by a search of the literature are limited to those that have been considered by previous researchers. Both *systematic reviews* and *meta-analysis* address these first two problems, and the third is met by *pooling observations* from multiple studies.

A *systematic review* is, in a nutshell, a more rigorous form of a literature review that attempts to exhaustively identify and incorporate all high quality published work on a topic (Higgins and Green 2011). These reviews clearly specify a protocol for searching databases and citation indices for relevant work, and they have clear criteria for which of the identified publications will be included in the review. Thus, they avoid some of the selection bias associated with standard literature reviews, which have less rigorous search and inclusion criteria.

Often, the results of the publications assembled in the systematic review will then be statistically combined to form *quantitative* summaries of the published works, which is a process called *meta-analysis*. The first step in producing such summaries is to identify, if possible, a common effect measure available in the collected studies. For example, if the goal of a review is to determine the relative effectiveness of buprenorphine and methadone for treating heroin addiction, then one effect measure might be the *odds ratio* comparing the odds of relapse into heroin use for subjects treated with buprenorphine to the odds of relapse for methadone. One complication is that the odds of

relapse may depend on how long the subjects are followed, and length of follow-up will likely vary between studies included in the review. However, as long as the *ratio* of the odds comparing the two treatments does not depend heavily on the duration of follow-up, it may still be sensible to use the odds ratio as the common effect measure.

The next step is to identify an estimate of the chosen effect measure, together with some measure of the precision of this estimate, such as sample size or standard error, from each of the papers included in the review (Deeks et al. 2011). Often at this point, a *forest plot* containing estimates and confidence intervals from each study is produced to check for outlying and influential studies. Finally, the estimates and their precisions are combined, usually with some form of precision-weighted average, to obtain the desired overall mean effect estimate across all studies. In addition, a measure of the uncertainty of this estimate, such as a confidence interval, is often calculated and reported, and an overall statistical hypothesis test may be conducted (DerSimonian and Laird 1986). The quantitative inferences produced from these individual study results allow us to reach stronger conclusions than would be possible with the qualitative summaries produced in classical literature reviews.

5.5.3.2 Pooling Subjects or Observations from Multiple Studies

While meta-analyses allow us to effectively reuse data from previous studies to quantitatively assess previously investigated hypotheses, they cannot be used to assess novel hypotheses that are not addressed in the published works included for review (Hussong et al. 2013). To address new scientific questions with data from multiple previous studies, the first step is to obtain the original, subject-level data from the individual studies (often called *individual patient data*, or *IPD)*, much alike single-study secondary analysis (Stewart et al. 2011; Stewart and Tierney 2002). The next step is to pool the observations from all studies into a single large sample and fit a single statistical model to the pooled sample data, which accounts for any *heterogeneity*, or differences, with respect to the populations and protocols used in the various studies (Riley et al. 2010).

Continuing the meta-analysis example, suppose a researcher wishes to apply a logistic regression model to a pooled sample to calculate the odds ratio comparing the odds of heroin relapse in buprenorphine subjects to methadone subjects. If, as mentioned before, follow-up was substantially longer or shorter for certain studies contributing data to the analysis, then study-specific random effects could be included, which would capture the higher or lower odds of relapse induced by the longer or shorter follow-up periods. Essentially, this would be treating studies as different clusters, as described in the section on Multilevel Modeling of Correlated Data. A more complicated problem arises if subjects not only differ between studies with respect to their overall average response level, but subjects also differ in their response to treatment. This *heterogeneity of treatment* effect must be modeled using treatment-by-study interactions, which will also usually be included as random effects.

5.5.4 Missing Data

Missing data is inevitable in practice, and this is especially true in secondary analysis. Missing data occur for different reasons. Subject dropout is a typical cause of missing data that occurs in longitudinal studies; in such studies, the measurement is repeated over time. Missing values arise when participants drop out before the study ends and then one or more measurements are missing. Omission is another cause of missing data; this occurs when participants refuse to respond to some questions in a questionnaire. Some of these missing values can be a result of study design because the questions asked are not applicable to all subjects. For example, the age of first heroin use is only applicable to heroin users and so non-heroin users would have no response to such a question. Missing data also occurs when data collection is done improperly or

mistakes are made in data entry. Of course, in secondary analysis, some nonresponses may be due to skip logic, so these can be recoded and not treated as missing values. For example, a nonresponse on the frequency of heroin use in the past 30 days can be considered as *no use* instead of a missing value for the participants who reported never using heroin.

5.5.4.1 Missing Mechanism

The most intuitive response to missing data is to delete observations with any missing values (also called *complete case analysis*). From the perspective of statistical theory, the effectiveness of the complete case analysis depends on the missing data mechanism. Rubin (1976) classified three types of missing data mechanism: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

To illustrate, let $y_i = b_0 + b_1 x_i + \epsilon_i$, and let y_i denote the number of days of heroin use at follow-up and x_i denote treatment participation, with $x_i = 1$ for *yes* and $x_i = 0$ for *no*. The unbiased estimate of b_0 is the average observed value of y_i among nonparticipants and the unbiased estimate of b_1 is the average observed value of y_i among participants minus the average observed value of y_i among nonparticipants. With MCAR data, the probability of having a missing value on y_i is not related to any variables (e.g., treatment) under study. Then the average observed values of y_i among participants and nonparticipants still unbiasedly reflect the true heroin use of each group, though less precisely and with less observed cases. Correspondingly, the estimates b_0 and b_1 based on these group means are still unbiased, though less precise. With MAR data, the probability of having a missing value on y_i is related to the observed values in the study (e.g., treatment) but is not related to the unobserved values of y_i. Suppose participants and nonparticipants now may have a different chance of having missing values. However, the average observed values of y_i among participants and nonparticipants still unbiasedly reflect the true heroin use of each group since the missingness is still random within each group. As a result, the b_0 and b_1 estimates based on these group means are unbiased too. In statistics, both MCAR and MAR are called *ignorable missing* because ignoring the missing values, as in the case noted above, would not cause any bias to the estimators of b_0 and b_1. With NMAR, the probability of having a missing value on y_i is related to the unobserved values of y_i. For example, more severe heroin users across groups might be less likely to provide responses. As a result, the average observed values of y_i among participants and nonparticipants both could underestimate the true heroin use of each group, and therefore, the corresponding estimates b_0 and b_1 based on these group means could be biased. In this situation, the missing values are not ignorable and NMAR is called a *nonignorable missing condition*.

5.5.4.2 Handling Missing Data

The commonly used complete case analysis usually requires MCAR. However, in the example above, it only requires ignorable missingness. Complete case analysis due to loss of cases would have larger standard errors, wider confidence intervals, and a loss of power in testing hypotheses. In addition to complete case analysis, some other heuristic methods of missing-value handling include imputation from a randomly selected similar record, from the last observation in an ordered dataset (or *hot-deck imputation*), replacing any missing value with the mean of that variable (*mean substitution*), or replacing any missing value with the predicted value of that variable based on the regression model from the complete cases (*regression substitution*). Though frequently used, these heuristic methods have less statistical rigor, and their validity often is questionable.

For missing values in secondary analysis, two popular handling methods with a strong statistical foundation and rigor are *expectation maximization algorithm,* or EM (Dempster et al. 1977), and *multiple imputation*, or MI (Rubin 1987; Schafer 1997). Both EM and MI require ignorable missingness for valid statistical estimation and inference. Given MCAR or MAR data, the EM algorithm estimate b_0 and b_1 in the above case through an iteration process. In each

iteration, the missing values of y_i are imputed from its posterior means, conditional on the observed values of y_i and x_i (E-step), and then the estimates b_0 and b_1 are calculated based on the imputed data without missing values for the next iteration (M-step). The EM algorithm has been applied to various models and situations in statistics over the past two decades (Little and Rubin 2002; Schafer 1997). However, despite its popularity and generality, the EM algorithm often is model- or situation-specific, and thus is difficult for people with a limited statistical background to implement. On the other hand, many statistical software packages (e.g., SAS, SPSS, Stata, R base) assume a complete case analysis and exclude the EM algorithm for missing-value handling in their commonly used analysis procedures. As a result, special software packages or programs that include the EM algorithm for missing values may be required, depending on specific analyses.

Compared to the EM algorithm, the MI method is less model- or situation-specific and has been implemented in many standard software packages (e.g., SAS, SPSS, Stata, Splus, R). The MI method often includes three steps. In the first step, a selected imputation method, such as the predictive mean matching method (Heitjan and Little 1991; Schenker and Taylor 1996), or the Markov chain Monte Carlo method, or MCMC (Schafer 1997) is used to create several replications with the missing values replaced by the imputed values. For example, with some specification (e.g., imputation method, the number of replications), SAS Proc MI (SAS Institute, Inc. 2013) can generate the replications. Then, in the second step, these replications are analyzed by the standard procedures in the same way as the complete data. In the above heroin-use example, the parameters b_0 and b_1 can be estimated for each replication by any software that includes linear regression. Then, in the third step, the estimates from each replication will be combined to calculate the final estimates of b_0 and b_1 and their standard errors. In practice, this final step can be conducted by some standard software, such as SAS Proc MIANALYZE (SAS Institute, Inc. 2013).

5.5.5 Propensity-Score Analysis

Assessing treatment effects or outcomes among different treatment conditions and/or groups is often of interest in the secondary analysis of observational data in the substance abuse field. Because observational data lack randomized assignment of participants into treatment conditions, observed differences in outcomes could be due to pretreatment differences (or selection biases) instead of treatment conditions, and therefore statistical procedures are needed to balance the data before assessing the treatment effects. Conventional methods in controlling for self-selection may involve pretreatment measures for the following reasons: (1) as to control measures or covariates in regression models (ANCOVA approach), (2) for matching, or (3) for stratification. When there are multiple control variables (i.e., a high-dimensionality problem), matching and stratification are more difficult to apply. If units in the treatment and control are balanced on a large number of covariates one at a time, large numbers of observations would be needed to overcome the "dimensionality problem," whereby the introduction of a new balancing covariate increases the minimum necessary number of observations in the sample geometrically. On the other hand, the validity of the ANCOVA approach rests upon the assumption that the mean outcome varies linearly with all confounders with identical slopes between treated and non-treated groups. Even though the inclusion of interactions between treatment and confounders may be helpful sometimes, the potential of violating the assumption grows as more confounders are added without interaction terms, and as the treated and non-treated groups become increasingly different with respect to these confounders (e.g., different degrees of nonlinearity between treatment groups). Therefore, it is more advantageous to use propensity-score analysis, which reduces multiple characteristics to a one-dimensional propensity score and matches the scores between treated and non-treated cases, and then use the matched sample to estimate treatment effects. By using a linear combination

of covariates for a single score, treatment and control groups are balanced on a large number of covariates. Increasingly, propensity-score analysis has become a popular statistical technique for reducing the impact of selection bias in estimation of causal effects using observational data (D'Agostino 1998; Feng et al. 2006; Rosenbaum and Rubin 1983; Rubin 1997; Rubin and Thomas 1996; Ye and Kaskutas 2008).

5.5.5.1 Procedures for Propensity-Score Analysis

While several propensity-score models are available, a propensity score is basically defined as the conditional probability of assignment to a particular treatment, given a vector of observed covariates (Rosenbaum and Rubin 1983). Propensity-score analysis involves the following steps. Step 1 involves the specification of a logistic regression model and the search for a best set of conditioning variables that optimizes estimates of the propensity scores. Step 2 is matching or resampling based on estimated propensity scores. Various propensity-score matching algorithms (e.g., greedy matching, Mahalanobis metric distance matching, both with or without propensity scores, optimal matching) have been developed using the propensity scores to match treated participants with the control participants, with the goal to make the two groups as much alike as possible in terms of estimated propensity scores. Because the common support region formed by the estimated propensity scores does not always cover the whole range of study participants, some treated participants may not have matched controls and some control participants may never be used; thus, propensity-score matching sometimes leads to a loss of study participants. Step 3 involves analysis based on the matched sample. Some propensity-score methods do not involve Step 2; for example, a propensity-score weighting approach omits propensity-score matching but conducts multivariate analysis using propensity scores as sampling weights. Propensity-score weighting avoids the problem of loss of sample participants, but can be sensitive to misspecification of the propensity-score model when estimated propensities are small (Kang and Schafer 2007).

An important component of any propensity-score analysis is to examine whether the propensity-score model has been adequately specified. This is often accomplished by examining whether the distribution of measured baseline covariates is similar between treated and untreated participants with the same estimated propensity score. If, after conditioning on the propensity score, systematic differences remain in baseline covariates between treated and untreated participants, this can be an indication that the propensity-score model has not been correctly specified. For example, with propensity-score matching, assessing whether the propensity-score model has been adequately specified involves comparing pretreatment characteristics of treated and untreated participants within the propensity-score matched sample.

5.5.5.2 An Empirical Example

The following is an example comparing long-term outcomes among women who were pregnant or parenting at admission to women-only (WO), versus mixed-gender (MG) programs participating in the California Treatment Outcome Project, or CalTOP (Hser et al. 2011). Of the intake sample of 4448 mothers, 3688 were in MG and 760 were in WO, and women in the WO programs generally demonstrated greater problem severity in many key life domains at intake. To assess the differences in outcomes between the WO and MG programs, we used propensity-matching procedures, as described below, to select an equivalent 500 women from each type of program.

The propensity score is a subject's probability of receiving WO or MG treatment, conditional on observed covariates for this analysis. Matched pairs were constructed using the nearest available Mahalanobis metric matching within calipers defined by the propensity score. Specifically, we considered WO programs as the "case" group and MG programs as the "control" group and performed the following procedures: (a) propensity score was computed based on 47 variables at intake, (b) women in both groups were first

stratified by treatment modality at intake and psychiatric severity at intake, (c) treatment modalities were residential, outpatient, and methadone maintenance, (d) psychiatric severity was indicated by the Addiction Severity Index (ASI) psychiatric score and separated using a median split as high and low categories, (e) all women in each group were divided into six substrata, (f) matched pairs were searched only among subjects that belonged to the same stratum, and (g) the matching procedure was divided into three steps and was started with pregnant women first.

Step 1—given each prestratified stratum, a matched subject from the MG programs was found for each pregnant woman in WO programs;
Step 2—of the remaining unmatched subjects, a matched subject from a WO program was found for each pregnant woman in a MG program, given each prestratified stratum;
Step 3—of the remaining unmatched subjects, a matched subject was found from a MG program for a subject in a WO program, given each prestratified stratum.

Applying propensity-matching procedures using a large number of pretreatment measures ($n = 47$), the baseline group differences were largely eliminated, but a few remained (e.g., ages of 30.5 vs. 31.9, pregnant status of 45% vs. 73%, and prior outpatient psychiatric treatment 33% vs. 41%). These differences may have influenced the long-term trajectories examined in the study. In addition, sensitivity analyses revealed significant differences in patient characteristics between the total and matched samples. It may be that the population of women at MG agencies does not sufficiently overlap the population of women at WO agencies to generalize results more broadly.

5.5.5.3 Additional Considerations

In the absence of an experimental design, the propensity score allows the design and analysis of an observational study to mimic some of the particular characteristics of a randomized controlled trial. Nevertheless, Rubin (1997) points out three limitations of propensity-score matching. First, propensity-score matching only accounts for observed (and observable) covariates and cannot control for unobserved selection bias. Any hidden bias due to latent variables may remain after matching because the procedure only controls for observed variables. Second, propensity-score matching works better in larger samples, with substantial overlap between treatment and control groups. Third, propensity-score matching does not handle a covariate that is related to treatment assignment but not related to the outcome, in the same way as a covariate with the same relation to treatment assignment but strongly related to outcome. Rubin recommended performing sensitivity analysis and testing different sets of conditioning variables to address the first limitation. It should be noted that propensity-score matching is a rapidly growing field of study and many new developments are still in a testing stage (Guo and Fraser 2010). Multiple methods for estimating program effects should be considered to increase confidence in the findings.

Acknowledgements The writing of this chapter was supported by the National Institute on Drug Abuse, Center for Advancing Longitudinal Drug Abuse Research (CALDAR, P30 DA016383, PI: Hser).

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.
Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.
Arbuckle, J. L. (2006). *Amos* (version 7.0) [computer program]. Chicago, IL: SPSS.
Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 2,* 238–246.
Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
Bentler, P. M. (2000–2008). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.
Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 719–725.
Biernacki, C., & Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics, 29,* 451–457.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach (Wiley series on probability and mathematical statistics)*. Hoboken, NJ: Wiley.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8,* 1–24.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13,* 195–212.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, lagrange multiplier, and Wald tests. *Multivariate Behavioral Research, 25,* 115–136.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design and statistical model. *Annual Review of Psychology, 57,* 505–528.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B, 34* (2), 187–220.

D'Agostino, R. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265–2281.

Dang, H. D. (2011). *A latent transition analysis of self-efficacy among men treated for cocaine dependence* (doctoral dissertation). **A**vailable from ProQuest dissertations and theses database (UMI No. 3472617).

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2011). Chapter 9: Analysing data and undertaking meta-analyses. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions, version 5.1.0 (updated March 2011)*. London, UK: The Cochrane Collaboration. Available from www.cochrane-handbook.org

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series. B, 39,* 1–38.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188.

Eliason, S. (1997). *The categorical data analysis system.* Version 4.0 of MLLSA. Iowa City, IA: University of Iowa.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

Feng, W., Jun, Y., & Xu, R. A (2006). *Method/macro based on propensity score and Mahalanobis distance to reduce bias in treatment comparison in observational study.* SAS Technical Report, paper PR05, pp. 1–11.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2010). *Fundamentals of clinical trials* (4th ed.). New York, NY: Springer.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

Green, S. B., Thompson, M. S., & Babyak, M. A. (1998). A Monte Carlo investigation of methods for controlling type I errors with specification searches in structural equation modeling. *Multivariate Behavioral Research, 33,* 365–384.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and application*. Thousand Oaks, CA: Sage Publications.

Heitjan, F., & Little, R. J. A. (1991). Multiple imputation for the fatal accident reporting system. *Applied Statistics, 40,* 13–29.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011).* London, UK: The Cochrane Collaboration. Available from www.cochrane-handbook.org

Homburg, C., & Dobartz, A. (1992). Covariance structure analysis via specification searches. *Statistical Papers, 33*(1), 119–142.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York, NY: Wiley.

Hser, Y.-I., Evans, E., Huang, Y., & Anglin, M. D. (2004). Relationship between drug treatment services, retention and outcomes. *Psychiatric Services, 55*(7), 767–774.

Hser, Y.-I., Evans, E., Huang, D., & Messina, N. (2011). Long-term outcomes among drug-dependent mothers treated in women-only versus mixed-gender programs. *Journal of Substance Abuse Treatment, 41*(2), 115–123.

Hser, Y.-I., Huang, D., Chou, C.-P., & Anglin, M. D. (2007). Trajectories of heroin addiction: Growth mixture modeling results based on a 33-year follow-up study. *Evaluation Review, 31*(6), 548–563.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9,* 61–89.

Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science, 22,* 523–539.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods and Research, 29,* 374–393.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.8 for Windows* [computer software]. Skokie, IL: Scientific Software International, Inc.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Hoboken, NJ: Wiley.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Hoboken, NJ: Springer.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A., & Collins, L. M. (2013). *PROC LCA and PROC LTA Users' guide* (version 1.3.0). University Park, PA: The Methodology Center, Penn State.

Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research, 46*(2), 266–302.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100,* 107–120.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Reviews in Psychology, 51,* 201–226.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490–504.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36,* 318–324.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., et al. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment, 9*(3), 199–213.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 369–377.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 345–368). Thousand Oaks, CA: Sage Publications.

Muthén, B., & Muthén, L. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U. S. National sample. *Journal of Studies on Alcohol, 61*(2), 290–300.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4,* 599–620.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463–469.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th Ed.). Los Angeles, CA: Muthén and Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric group-based approach. *Psychological Methods, 4*(2), 139–157.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods, 6,* 18–34.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal, 19,* 204–226.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal, 340*(7745), 521–525.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127,* 757–763.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52,* 249–264.

SAS Institute Inc. (2013). *SAS/STAT® 13.1 user's guide.* Cary, NC: SAS Institute Inc.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 proceedings of the Business and Economics Sections* (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman and Hall.

Schenker, N., & Taylor, J. M. G. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis, 22,* 425–446.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., et al. (2011). Conducting high-value secondary dataset analysis: an introductory guide and resources. *Journal of General Internal Medicine, 26*(8), 920–929.

Sörbom, D. (1989). Model modification. *Psychometrika, 54,* 371–384.

Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors.* Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.

Stewart, L. A., & Tierney, J. F. (2002). To IPD or not to IPD?: Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions, 25*(1), 76–97.

Stewart, L. A., Tierney, J. F., & Clarke, M. (2011). Reviews of individual patient data. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (version 5.1.0) [updated March 2011]. London, UK: The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Vermunt, J. K. (1997). *LEM 1.0: A general program for the analysis of categorical data*. Tilburg, NL: Tilburg University.

Vermunt, J. K. (2004). Latent Markov Model. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The sage encyclopedia of social science research methods* (pp. 553–554). Thousand Oaks, CA: Sage Publications.

Vermunt, J. K., & Magidson, J. (2013). *Latent GOLD 5.0 upgrade manual.* Belmont, MA: Statistical Innovations Inc.

Von Davier, M. (1997). WINMIRA program description and recent enhancements. *Methods of Psychological Research Online, 2,* 25–28.

Weiss, R. E. (2005). *Modeling longitudinal data*. New York, NY: Springer.

Weston, R., & Gore, P. A., Jr. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34,* 719–751.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology, 61*(6), 952–965.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics and Data Analysis, 50,* 1090–1104.

Ye, Y., & Kaskutas, L. A. (2008). Using propensity scores to adjust for bias when assessing the effectiveness of Alcoholics anonymous in observational studies. *Drug and Alcohol Dependence, 104,* 56–64.

Longitudinal Methods in Substance Use Research

6

Brent Teasdale and Jerreed Ivanich

6.1 Introduction

Substance use researchers utilize a variety of methods, a good deal of which involves the collection and analysis of survey data. Survey researchers, in turn, typically utilize either cross-sectional or longitudinal methodological approaches. In cross-sectional research, subjects are studied at only one point in time. The focus of this chapter will be on longitudinal research, which is defined as repeated observations across time. This chapter will review basic types of longitudinal research utilized in studies of substance use and abuse, including major longitudinal data collection efforts, different kinds of data analytic techniques for the analysis of longitudinal data, and concludes with a summary of the promise of longitudinal methods for substance abuse researchers.

Longitudinal studies can be classified into two types. First there are repeated cross-sectional studies. These studies typically survey different samples at each time period, using the same survey instrument. Because of their use of different samples, these studies do not allow the researcher to compare the same unit of observation across time, but they do allow the researcher to track trends across time in the variables measured, at the aggregate level. An alternative to the repeated cross-sectional study is the panel study. Panel studies survey the same unit of observation repeatedly across time. These studies allow the researcher to make within-person comparisons—comparing the same person to himself or herself over multiple time periods. We will return to each of these types in greater detail below.

6.2 Examples of Longitudinal Data Collections

As noted, repeated cross-sectional studies allow researchers to track trends in the variables measured over time; they do so by using the same measures on different samples surveyed at different times. One example of this type of research in the area of substance use is the Monitoring the Future Study (MTF). The MTF study is an ongoing study of American high school students conducted by the Institute for Social Research at the University of Michigan. MTF began in 1975 and has surveyed 12th graders every year since. In 1991, the MTF also began surveying 8th and 10th graders. Each year, approximately 50,000 8th, 10th, and 12th graders in approximately 420 public and private high schools are surveyed (www.monitoringthefuture.org). The MTF study has been influential in demonstrating the rates of substance use and abuse among American high school students, as

B. Teasdale (✉)
Department of Criminal Justice and Criminology, Georgia State University, Atlanta, GA, USA
e-mail: bteasdale@gsu.edu

J. Ivanich
Department of Sociology, University of Nebraska-Lincoln, Lincoln, NE, USA
e-mail: jerreedivanich@gmail.com

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_6

well as trends in attitudes and behaviors related to substance use across time. The study is so influential that the investigators have been called upon to testify before congress over a dozen times (www.monitoringthefuture.org).

Because repeated cross-sectional studies track the same variables over time, they can be utilized to follow trends over time. The MTF study has been tracking trends in drug use since 1975, enabling researchers to understand the trends in drug use, attitudes, etc. For example, Terry-McElrath et al. (2013), utilizing data from nationally representative cross-sectional samples of 12th grade students surveyed in the MTF project, tracked trends from 1976 to 2011 in the simultaneous use of alcohol and marijuana (SAM). Although high, results suggested a relatively flat trend in the simultaneous use of alcohol and marijuana,[1] with a slight decrease over time. The study's design revealed consistencies in reasons given for SAM use over time, which included certain social contexts of marijuana or alcohol use (e.g., park/beach, car, party), which may be used to inform tailored prevention efforts.

Another example of a repeated cross-sectional design is the National Survey on Drug Use and Health (NSDUH), formerly known as the National Household Survey on Drug Abuse, which is the primary source of data on drug use for the noninstitutionalized U.S. population aged 12 years and older (Substance Abuse and Mental Health Services Administration 2013). Data has been collected since 1971, making the NSDUH the largest longitudinal study of drug use in the U.S.; providing important estimates of the trends in use behaviors and how those behaviors vary by key demographics (gender, race, age, etc.). In addition, information regarding treatment is also collected, providing key insights on the need for and use of treatment services.[2] Recent results from the 2013 NSDUH report suggested that 8.3% of respondents used illicit drugs in the past month based on the 2002 survey, a figure that increased to 9.2% by 2012. Furthermore, 6.2% of respondents reported using marijuana in the past month based on the 2002 survey, while 7.3% reported its use in 2012. In contrast, use of psychotherapeutics remained relatively constant across the period (past month prevalence of 2.7% in 2002 and 2.6% in 2012). Cocaine use trended down over the period from 2002–2012, with 0.9% using in 2002 and 0.6% using in 2012 (both past month prevalence).

In contrast to the repeated cross-sectional designs, panel designs follow the same subjects across time. These data collection efforts allow the researcher to track an individual's behavior over time, rather than aggregate trends in the variables, which are produced from repeated cross-sectional studies. For example, the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) is a two-wave panel study funded by the National Institute on Alcohol Abuse and Alcoholism. The first wave of data was collected in 2001–2002 and the second wave was collected in 2004–2005. The data collected included all aspects of alcohol use, including initiation, service utilization, comorbidity with drug and mental health problems, and underage drinking. Analyses of the NESARC have investigated a variety of topics including gender differences in anxiety disorders (Vesga-Lopez et al. 2008) and alcohol consumption among young adults (Chen et al. 2004). An interesting longitudinal use of the data was presented by Lopez-Quintero et al. (2011) in which they studied ethnic group differences in the transition from use to dependence using the two waves of the NESARC. They found that the probability of transitioning from use to

[1]Students who reported any past 12-month marijuana use were asked: "How many of the times when you used marijuana or hashish during the last year did you use it along with alcohol—that is, so that their effects overlapped?" In the analyses, any SAM use decreased from a high of 74% in 1980–82 to 62% in 2011. SAM use "most or every time" remained generally stable at around 20% through the mid-1990s, but then decreased significantly through 2011.

[2]In 1999, the NSDUH began using computer-assisted interviewing. Consequently, caution is necessary when estimating trends in drug use before and after the methodological change, as their measurement may not be consistent.

dependence[3] was 67.5% for nicotine users and 22.7% for alcohol users. Furthermore, they found that there were significant differences in the probability of transitioning from use to dependence by racial and ethnic groups. Specifically, Native Americans or Alaskan Natives had higher probabilities than White Americans of transitioning to dependence across all of the substances investigated. This information, coupled with identification in this study of several common predictors of transition dependence (e.g., history of any mental disorder), highlighted the importance of continued outreach and treatment for these populations.

6.3 Longitudinal Epidemiological Research on Substance Use

Several theories have impacted the longitudinal study of substance abuse. Specifically, theories of alcohol use trajectories and the dominant theories explaining stage-specific development are the driving inquires within longitudinal substance abuse research. To study these topics, researchers have often used longitudinal techniques as a tool to uncover trends in substance abuse, comparisons between groups, and the key factors that trigger one's progression of substance abuse through developmental stages.

6.3.1 Alcohol Trajectories

Although substance abuse is often associated with illegal drugs, researchers have often noted the importance of reviewing alcohol use separately from more illicit drugs, as alcohol may have different driving factors for inception and continued use than do illegal drugs (Chassin et al. 2004). In the U.S., alcohol is not illegal for consumption by individuals older than 21 and is much more accessible than other illicit drugs, making it more susceptible to exposure and abuse for all. The benefit of reviewing alcohol-related abuse separate from other substances in longitudinal studies provides a partition between other illicit drug abuse factors and alcohol abuse characteristics. The added clarity allows researchers and practitioners to estimate trends in use and the impacts of variables that may be unique to each substance.

Adolescents have been cited as the most at-risk population to be impacted by alcohol. Consequently, much research has focused on adolescent attitudes toward alcohol use (Bates and Labouvie 1997). Furthermore, the developmental process in adolesence becomes a primary focus in alcohol abuse research. This area of research has been justified because of the predisposed sensitivity that becomes so pivotal in which trajectory of future alcohol use one is placed. Having information that tracks an individual's drinking behaviors combined with other factors over time allows for insight into the different trajectories of drinking behaviors and also what factors are more influential within each different trajectory group.

Longitudinal research has focused primarily on the study of how one's individual environment and social factors contribute to drinking behavior. Several elements emerge in the field as to what may be causing adolescents to begin drinking and what other factors may cause them to persist in their drinking. Common variables studied overtime to understand the onset factors associated with youth drinking include gender, parent alcoholism, friends' participation in the use/abuse of alcohol, parental divorce, and delinquency (Hawkins et al. 1992). The variables listed have found continued support, as contemporary studies still find them to be important risk factors associated with alcohol use and dependency (Chassin et al. 1996, 2004). Findings further indicate that individuals can be placed in trajectories that uncover patterns of alcohol use and dependency over time (Chassin et al. 2004). Alcohol trajectories in adolescence are typically grouped by the dependent variable (alcohol use) and researchers study the independent variables

[3]NESARC surveys included extensive questions covering the DSM-IV criteria for alcohol- and drug-specific abuse and dependence for 10 classes of substances (sedatives, tranquilizers, painkillers, stimulants, cannabis, cocaine or crack, hallucinogens, Inhalants/solvents, Heroin, alcohol, and nicotine).

that might place an individual in a given trajectory group. Studies are often divided between two primary foci based on the common trajectories identified. The first and primary focus has explored trajectories of adolescent development into adulthood (Chassin et al. 2002, 2004; Maggs and Schulenberg 2004). The second major area of alcohol trajectories that is often studied is the trajectories of adult substance abuse (Curran et al. 1998; Hussong et al. 2001).

Individual trajectories in alcohol use vary based on several factors. Primarily, studies examine the likelihood or extent to which an individual will abuse alcohol in the future based on factors such as age at onset of drinking, amount of drinking at different developmental stages of life, friends' alcohol use, parent drinking, stressful life events, and social attitudes toward alcohol. Additionally, common factors that predict trajectory membership include gender, parental alcoholism, psychological disorders, and culture (Hawkins et al. 1992). For example, Chassin and associates (2004) outlined adolescent trajectories of alcohol use and abuse starting at the age of 11 and ending when participants were 30. The trajectories were reviewed in three separate groups, light alcohol and drug use, moderate and experimental alcohol and drug use, and heavy alcohol and drug use. In each trajectory, a peak of alcohol consumption across all three groups was found in the age range of 23–26. Additionally, all three groups began to desist after the age of 26. However, the amount of alcohol consumed early in one's life was a significant factor in determining one's future drug and alcohol use (Chassin et al. 2004). These findings are consistent with the common notion that the age at which an individual begins using alcohol is a significant predictor of future use and/or abuse (Bates and Labouvie 1997; Chassin et al. 2002).

According to the Chassin et al. (2004) study, only 11.3% of individuals were found to be lifelong abstainers from alcohol and drug use. This is not to suggest that the majority of individuals face dependence on alcohol or drugs, as it was found that 61% never report dependence on alcohol or drugs. The most at-risk population for dependency was the heavy use group, with 80% reporting dependency. This heavy use group also exemplified many life factors that have been found to be associated with heavy drinking—parent alcoholism, onset of drinking early in life, parental divorce, and friends' alcohol use (Chassin et al. 2004).

In one classic example of longitudinal theorizing, Bates and Labouvie's (1997) study of 1257 Rutgers Health and Human Development Project participants gathered information from individuals aged 12, 15, and 18 years at time one, through telephone recruitment in New Jersey. The first follow-up occurred approximately three years after enrollment in the study. The third data collection occurred 3 years after the second follow-up and the final retest was conducted 7 years after the third retest yielding ages of 25, 28, and 31 for participants in the final data collection. Bates and Labouvie (1997) utilized 16 measures of risk factors in their study.[4] These measures include: cognitive structure, harm avoidance, religiosity, academic performance, impulsivity, Personality Research Form (PRF) play, disinhibition, experience seeking, self-derogation, emotional outbursts, deviant coping, delinquency, negative activities with friends, sibling deviance, friends' deviance, and stressful life events.

The authors' efforts yielded key insights as to what factors may place adolescents at the greatest risk for drinking. Results indicated that individuals whom had heavy use at early ages were more likely to continue alcohol use, but only when certain environmental and social factors were present. Heavy drinking patterns at age 18 had a low to moderate correlation of $0.22 < r < 0.32$ for males; $0.34 < r < 0.40$ for

[4]The study identified three subsets of individuals representing prototypical use trajectories of (1) consistently low alcohol and drug use during adolescence and early adulthood; (2) heavier alcohol and drug use in adolescence compared to early adulthood; and (3) persistent heavier alcohol/drug use through transition to early adulthood; all differing on a number of risk factors for persistent use into adulthood.

females with future use (age 21–23), when controlling for other risk factors (Bates and Labovie 1997). Bates and Labouvie concluded that their study supported the notion that younger age at onset was associated with persistent drinking problems, but the focus should be placed on risk factors in the individual's life beyond simply age at onset of alcohol use (1997). Bates and Labouvie's study pointed to a much-needed direction—the utilization of longitudinal studies over longer periods of time to better understand the relationships of environmental and intrapersonal risk factors with alcohol use (Bates and Labouvie 1997). Having data that covers longer time frames may also add to the understanding of developmental stages that alcohol and other drugs may have in an individual's growth.

6.3.2 Gateway Hypothesis

The gateway hypothesis is one of the most cited stage-developmental theories of substance abuse. The gateway hypothesis draws from the concept that certain drugs will act as a starting point and pathway into more illicit drug using behaviors. Kandel's (1975) seminal piece observed New York high school students using a series of three self-reported surveys, covering the span of 2 years. The pathways or "stages" found by Kandel (1975) included a four-step process. At stage one, an individual is either a nonuser or user of legal drugs (beer/wine). Stage two begins the transition from legal drugs to more illicit drugs, those that had been involved in the use of beer and wine now are users of cigarettes or hard alcohol. The third stage follows the use of cigarettes or hard alcohol and then feeds into the use of cannabis. The fourth and final stage is a progression from cannabis to hard drugs (pills, LSD, heroin, tranquilizers, etc.). The data suggests that substance abuse is not as straight forward as one's availability or desire, but is a gradual progressive process that allows individuals adjustment periods to move from one stage of use to increasingly more illicit drug use patterns, where each previous stage acts as a gateway to the next. Conversely, she argues that drug users do not stop using directly, but slow down and regress to previous stages as part of the process of becoming nonusers. Lastly, Kandel (1975) acknowledges that substance use in one stage does not cause the user to transition to the next stage, since many individuals do not advance to the next stage, but rather acts as a gateway to the next.

Kandel concluded that more research needed to explore the stability of risk factors over time, as well as identify potential protective factors compensating for early risk in developmental patterns of use. Indeed, the gateway hypothesis has sparked a good deal of interest, leading to a large number of studies that have evaluated different aspects of the hypothesis. Although many of Kandel's original findings still find support, different approaches are being taken to explain the reasons for each stage. Since the inception of the gateway hypothesis, the majority of the work being done has utilized self-reported surveys to solicit information from adolescents. The use of animal models, chiefly rats, is an emerging method in longitudinal research in evaluating the gateway hypothesis. Animal models present a unique possibility for studying causal factors in onset and persistence, which other methods would not allow (Grunberg and Faraday 2002). Animal models have shown that exposure to substances during adolescence is more likely to lead to continued use than is first exposure during adulthood (Levin et al. 2003). Additionally, animal studies are pushing the preconceived notions of their limitations by testing social/behavioral (e.g., peer pressure) traits (de Bono and Bargmann 1998). Overcoming the commonly perceived limitations of animal models, that they cannot evaluate social behavior, will advance their contributions in understanding the gateway hypothesis, by including factors such as peer pressure, strain, stress, family relationships, etc.

In the emerging studies being conducted to test the gateway hypothesis, life course methods are currently being employed. Life course

methods focus on the changes in the outcome behavior and connections to social and personal climates over a long period of time, to better account for the factors that may be causing an individual to exhibit a set of behaviors (e.g., substance abuse). These methods are valuable for their ability to find distinctive patterns that lead individuals to use illegal substances, identify critical events that are important predictors of substance abuse, and for finding the factors that contribute to persistence in substance abuse and those factors that are implicated in desistance from drug use (Hser et al. 2007).

Fergusson et al. (2006) conducted a study that employed a life course method in testing the gateway hypothesis. Their study used data collected on 1265 New Zealand individuals from birth until the age of 25. This unique sample of individuals allowed researchers to conduct a more in-depth analysis of factors often omitted from common survey data. Reviewing individuals across a large swath of their life better enabled the researchers to find events or patterns that may make an individual more likely to use illegal substances, than cross-sectional surveys. Fergusson and colleagues found strong support for the gateway hypothesis, as individuals that had been involved in cannabis drug use in adolescence progressed into illicit drug use in early adulthood. In particular, Fergusson and colleagues found that by age 25, 42% of individuals used illicit drugs outside of cannabis. Of the cohort that had used drugs outside of cannabis, 82% reported the use of hallucinogens (ecstasy, LSD) (Fergusson et al. 2006).

The phenomenon of substance abuse is clearly complex; researchers that utilize longitudinal methods have a distinct advantage over other methodological approaches. Tracking individuals over a period of time provides more clarity to the overall picture of substance abuse by better controlling for the time ordering of events and better understanding long-term impacts of use behavior on individual outcomes. Due to the inherent instability that is associated with substance abusers, tracking their behaviors over time has added, and will continue to add, to our knowledge base.

6.4 Examples of Longitudinal Research in Substance Abuse Prevention

Longitudinal evaluations have been consistently utilized in the area of substance abuse prevention. For example, Drug Abuse Resistance Education (D.A.R.E.), one of the most widely publicized programs, has been subjected to a number of longitudinal evaluations. Lynam et al. (1999) evaluated D.A.R.E. 10 years after administration, as a follow-up to the initial 5-year evaluation conducted by Clayton et al. (1996).[5] Overall, in their 10-year follow-up study of 1002 individuals who received D.A.R.E. (or "a standard drug education curriculum") in the 6th grade, they found that D.A.R.E. did not significantly impact alcohol, tobacco, marijuana, or illicit drug use, 10 years after administration (590). The strength of the findings from this longitudinal evaluation, if replicated, may call into question the value of an expensive program like DARE when compared to usual instruction (in the control conditions teachers did their usual drug education program as part of their standard heath curriculum) in preventing drug and alcohol abuse. Potential alternatives include a substantive redesign of either the curriculum or delivery mode as potential solutions.

Another example of a longitudinal evaluation of a substance abuse prevention curriculum delivered by D.A.R.E. officers was The Adolescent Substance Abuse Prevention Study (ASAPS), which was a project funded by the Robert Wood Johnson Foundation to analyze the effects of a substance abuse prevention curriculum known as Take Charge of Your Life (TCYL) using a prospective, cluster-randomized

[5]Both the 1999 and 1996 analyses utilize data from the same longitudinal evaluation of DARE that began in September 1987 with a 1987–88 6th grade cohort in a Midwestern metropolitan area with a population of 230,000. As can be expected, the major design and methodological confound is attrition over the lengthy follow-up period. However, analyses of attrition by condition at each follow-up showed little effect on the results, with missing participants more likely to be older males who reported using cigarettes in the 6th grade (Lynam et al. 1999).

experimental design (Sloboda et al. 2009). The ASAPS followed approximately 17,000 adolescents from the 7th through the 11th grades, including preventative interventions in both the 7th and 9th grades. Study schools came from six metropolitan regions and from schools within a 50-mile radius of each city center. These areas included: Detroit, MI, Houston, TX, Los Angeles, CA, Newark, NJ, New Orleans, LA, and St. Louis, MO. TCYL was delivered by trained Drug Abuse Resistance Education (DARE) officers. Research from the ASAPS suggests that DARE officers delivered the program with high fidelity (Sloboda et al. 2009). Outcome analyses suggested that adolescents who did not use alcohol and tobacco in the 7th grade were more likely to initiate use by the 11th grade in the treatment condition; however, 7th grade marijuana users were more likely to desist from use by the 11th grade if they received the treatment.

Another example is the major analysis of the Project Alert substance abuse prevention program, which consisted of a longitudinal data collection and analysis. Ellickson and Bell (1990) randomly assigned 30 schools to three conditions—a control condition that did not deliver Project Alert, a treatment condition that consisted of an adult-led Project Alert curriculum, and a treatment condition consisting of a peer-led Project Alert curriculum.[6] The longitudinal study followed the 6527 students, spread across 8 communities in California and Oregon, through four waves of data (two in the 7th grade and two in the 8th grade). Results suggested that program impacts on alcohol were across all user types and decayed quickly. That is, the treatment reduced use among baseline nonusers, experimenters, and users during the 7th grade, but impacts were not significant by the start of 8th grade. Results differed for cigarette use. Impacts were typically not seen until after the booster session and were confined to the experimenter and users groups. The program did not help previously confirmed smokers.[7]

In a similar vein, Pentz et al. (1989) conducted a longitudinal evaluation of the Midwestern Prevention Project. The evaluation utilized a 2×2 design where students were randomly assigned to treatment or comparison group and the program was either delivered in the 6th or 7th grade in two locations: Kansas City and Indianapolis. The participants were followed for 2 years after the intervention. Outcomes measured at time 2 suggested significant reductions in past month, past week, and past day smoking behaviors.

Lastly, one of the more researched programs is the Good Behavior Game (GBG) (Embry 2002). The intervention was designed to socialize children to the student role and reduce disruptive behaviors, including substance abuse disorders. In one example of a longitudinal evaluation of the preventive impacts of the GBG, Poduska et al. (2008) estimated generalized linear mixed models and generalized additive models to evaluate the program's impact on service utilization for mental health, emotional, behavioral, and drug-related problems. They found significant reductions in adult service utilization among males who received the treatment, but not females. Another longitudinal evaluation of the GBG, conducted by Kellam et al. (2008, 2011, 2014) found that administration of the GBG had significant impacts on lifetime diagnoses of substance abuse/dependence disorders in young adulthood,[8] demonstrating lasting impacts of the program. Impacts were highest among males already exhibiting high-risk behavior during the period of intervention. A replication of the GBG in Oregon that followed students from 5th grade to the end of high school showed significantly reduced use of tobacco, alcohol, and drugs (DeGarmo et al. 2009).

[6]The Project ALERT curriculum was developed using the social influence theoretical model.

[7]This was not totally unexpected, as Project ALERT was designed to keep nonusers from becoming involved with drugs and both nonusers and experimenters from making the critical transition to user. It was not designed for committed users.

[8]Subjects were followed up at multiple time periods. The evaluation used a modified version of the Composite International Diagnostic Interview (CIDI) designed to reflect the DSM-IV criteria for determining lifetime, past year, and past month occurrence of substance abuse and dependence disorders.

6.5 Examples of Longitudinal Research in Substance Abuse Treatment

Substance abuse *treatment* studies answer a different question than a majority of substance abuse related studies. Much substance abuse research concerns itself with questions related to the onset, maintenance, and causes for desistance of individuals. Substance abuse treatment research questions review how effective treatments have been. Additionally, substance abuse treatment research addresses what causes individuals to become "career treatment" individuals. A majority of treatment studies are conducted utilizing longitudinal methods, for their ability to collect data on one specific individual or program over time. Studying individuals or programs over time allows researchers to analyze trends of treatment utilization and key factors that create a successful treatment. Below, we discuss major treatment programs, coupled with their respective goals and outcomes.

6.5.1 Treatment Programs

One of the largest programs that was established for evaluating and studying substance abuse treatment programs was the Drug Abuse Treatment Outcome Study (DATOS). The National Institute on Drug Abuse (NIDA) established the efforts of DATOS in 1990. In 1995, four other research centers were funded to pursue coordinated, yet independent research based on DATOS (DATOS—About DATOS 2007). The four research centers include:

- The National Development and Research Institutes (NDRI) at North Carolina
- Texas Christian University (TCU) in Fort Worth
- The University of California at Los Angeles (UCLA)
- The NIDA Services Research Branch.

Each research center has since focused on a specific area of substance abuse treatment research. The goal of the institutions was, "to advance scientific knowledge about the effectiveness of drug abuse treatment as it is typically delivered in the United States."[9]

Prior research conducted before DATOS generally originated from two major outcome studies, the Drug Abuse Reporting Program (DARP) and the Treatment Outcome Prospective Study (TOPS). DARP was a primary mechanism for collecting treatment and substance abuse data in the 1970s and TOPS in the 1980s.[10] Both studies informed the direction and research questions of DATOS and in combination with DATOS they have given researchers and the government a rather comprehensive picture of the trends and patterns of substance abuse treatment in a longitudinal format, with over 30 years of information (http://www.datos.org/background.html). The advantages of this large longitudinal collection of data have extended into the scholarly field and the development of empirical public policy. The collection of data has allowed researchers to analyze trends of treatment outcomes, which has given practitioners and society the ability to refocus on the problem areas noted in the data. Additionally, DATOS has produced a large number of scholarly publications that add to the drug treatment knowledge base.

[9]For additional information on DATOS, see http://www.datos.org/aboutdatos.html. Also see Hubbard et al. (1997) for a detailed description of DATOS.

[10]DARP collected data from clients admitted to federally funded treatment agencies between 1969 and 1972. Data on substance abuse were collected at intake, during treatment, and at a series of follow-ups measuring outcomes up to 12 years post treatment. The TOPS study was a longitudinal, prospective cohort design that collected information on clients of treatment programs in 10 cities between 1979 and 81. Follow-up data included interviews 1 and 2 years after treatment with clients admitted in 1979; follow-ups 90 days and 1 year after treatment of clients who entered treatment in 1980; and follow-ups 3–5 years after treatment of clients who entered programs in 1981.

6.5.2 Longitudinal Studies in Substance Abuse Treatment

Researchers that are concerned with the effectiveness of treatment programs typically utilize longitudinal, follow-up studies. These follow-up studies typically observe the individuals that have received treatment about 6 months after program completion. Follow-up interviews and questionnaires gather information on the individual's substance use since the time of treatment along with other questions that may be relevant to the researcher's agenda (i.e. criminal behavior, moving patterns, family connections, etc.). It is with these follow-ups that programs are often evaluated (McKay and Weiss 2001). Follow-up studies have a mixed approach to the time and amount of follow-ups that occur. It is not uncommon for research to take one single follow-up several months after an individual has ended treatment. Researchers are becoming more aware of the complex dynamics that are involved in substance abuse treatments; therefore, a call for more frequent follow-ups over a longer period has been sounded in recent years (McKay and Weiss 2001). This will help researchers in their ability to form a complete picture of what factors may cause an individual to relapse, what substances are more at risk, and at what point in time after treatment.

Two forms of follow-up studies are found in treatment research. The first type of follow-up design is a one-time follow-up, where a certain time frame has elapsed from the treatment and contact is made with the individual. For example, Hubbard et al. (1997) used data collected from 2966 clients in the DATOS. The data was collected from a one-time follow-up interview 12 months after treatment. A key finding from the project was that, the odds of consuming cocaine on a weekly basis were lowered when the individual spent a longer period of time in treatment (Hubbard et al. 1997).

The benefit of the longitudinal design was in understanding which type of treatment is most effective (a critical concern to practitioners and policymakers). Hubbard et al. (1997) found that individuals in long-term residential, short-term inpatient, and outpatient drug-free programs reported a 50% reduction in daily and weekly use of cocaine 1 year after treatment, compared to the year before treatment (Hubbard et al. 1997). Long-term residential patients that spent 6 months or longer in treatment saw a 50% decrease in illegal activity and 10% increase in full-time employment (Hubbard et al. 1997). The findings of Hubbard and associates contributed to our awareness of the impact different types of treatment stays have.

The second form of follow-up used by researchers is less common, but perhaps more informative. This type involves multiple follow-ups. The benefit of utilizing multiple follow-ups is the ability to gather data over longer periods of time, for a richer understanding of treatment outcomes. For instance, Simpson et al. (2002) made use of a multiple follow-up design to study the idea that individuals that have completed treatment but have been involved more recently in drug consumption have increased odds of continued drug use. The authors used data from the DATOS. The researchers reviewed the cocaine-dependent subsample within DATOS and their 1-year follow-up interviews, coupled with their 5-year follow-up interviews. Utilizing a longitudinal design by following up with recently released individuals from treatment faculties, they were able to test drug consumption rates at different points in time after individuals were released from treatment centers. Findings indicated that 17% of individuals (at the 5-year follow-up period) had entered into a treatment facility, and of these individuals their likelihood of personal severity index scores (measurement of seven common drug-related activities) were much higher than individuals not found in treatment centers. Individuals with higher personal severity index scores showed the most significant improvement at the 5-year follow-up. Specifically, the percentage reporting weekly heroin use dropped from 27 to 7%. In addition to decreased heroin use, compared with the time prior to original intake, these individuals had significantly decreased in areas of daily alcohol use (49% to 14%), recent arrest (53% to 28%), and psychiatric

symptoms (from 92 to 44%) (Simpson et al. 2002). The use of multiple follow-ups lent itself well to answering this particular question because data collected at multiple points in time revealed how treatment can be beneficial to certain groups after a longer period of time, particularly those at highest risk. In another example, Hubbard et al. (2003) expanded on previous work by using DATOS to examine both the initial 1-year as well as 5-year drug treatment outcomes, with results illustrating the overall stability of substance abuse treatment across a number of outcomes.

Another topic commonly studied in the field of substance abuse treatment is the study of career treatment individuals. Career treatment refers to individuals that form a habit of returning to treatment facilities. These individuals are of interest to researchers because they are unique cases that indicate a sense of dependency or lack of treatment effectiveness (Anglin et al. 1997). Career treatment individuals no longer think of treatment episodes as the end to their problems, but part of a cycle that cannot be broken. Research that studies these concerns utilizing longitudinal methods reviews the trends of successes and failures of the individuals, in hopes of rectifying the problem of returning to treatment. More specifically, they are able to use longitudinal data to explore the complexities of addiction patterns in a manner that more effectively represents the chronic, relapsing nature of the condition, including its antecedents and consequences (Anglin et al. 1997; Hser et al. 1997). Possible solutions to the career treatment problem are also explored, including, but are not limited to, longer initial treatment programs, more intensive background screenings for individualized program development, tailoring treatment modalities for certain drug dependency patterns, and specialized programs for returning treatment seekers (Anglin et al. 1997; Hser et al. 1997). The efforts that are placed on these individuals may have benefits in the long term by reducing service utilization by high-risk individuals (Hubbard et al. 1997).

The DATOS has found application in the review of treatment career research, as well as treatment effectiveness. For instance, in addition to the work by Anglin and colleagues noted above, Grella and Joshi (1999) reviewed data from 7652 individuals from DATOS to examine gender differences for repeat treatment participants (career treatment individuals). The researchers hypothesized that the reasons that draw an individual to repeat treatment would be different for men and women. To assess these possibilities, individual data at intake to facilities were reviewed, as well as at follow-up 1, 2, 6, and 12 months later. Specifically, Grella and Joshi (1999) compared demographic and background information, addiction characteristics, treatment career characteristics, family relationships, criminal status, and mental health status between men and women as potential factors in treatment recidivism. Results of their work suggested that there is some gender specificity in the risk factors for long treatment careers. Specifically, men were driven to seek repetitious treatment for factors including family opposition to drug use and support of treatment. However, women cited different reasons for reaching out multiple times for treatment, including antisocial personality disorder and self-initiation (Grella and Joshi 1999).

6.6 Data Analysis Issues

The analysis of panel data allows the researcher to ask different kinds of questions and consequently requires different analytical techniques than the analysis of cross-sectional survey data. In this section, we will explore several different types of longitudinal data analysis methods. First, we review survival analysis. Next, we review the multilevel model for change and the possibilities of within-person analysis. Finally, we review group-based trajectory modeling.

Longitudinal data allows for the analysis of time until failure models. For example, a researcher might be interested in how long a person is in recovery before they relapse and what variables correlate with the length of time until an individual relapses. These questions require analytic techniques that account for the distribution of the outcome variable (time until

an event). This is particularly the case, when we study a group of people, not all of whom have experienced the event in question. For those individuals that have not experienced the event, they are what we call right-censored. That is, their score on the time until relapse variable is not known, because they have not yet relapsed. This is differentiated from left-censoring, which occurs when the onset period is unknown, so that total time at risk is not measured for that individual. To account for this type of censored outcome variable, a specialized set of techniques called proportional hazards models have been developed. These models, such as Cox regression are estimable in most statistical packages, such as SPSS, SAS, Stata, and Mplus. For an extended discussion of these methods, see Singer and Willett (2003) and Allison (1984).

Another possible question a researcher might ask regards the developmental course or trajectory of substance use over a given time period. In order to answer that question, the researcher would need to analyze slopes as outcomes. This kind of research investigates the shape of the time trend on the outcome. That is, the impact of time on the outcome is of primary interest. The effect of time is typically modeled via the use of the multilevel model for change (Singer and Willet 2003). In this case, waves of data are nested within people. The multiple observations per person cause a statistical problem known as dependence. That is, since the multiple observations are not independent (they all come from the same person), they tend to be correlated. This correlation violates the residual independence assumptions of single level models, such as OLS. This is resolved in the multilevel model by the inclusion of a second-level error term (typically denoted by μ_0), which captures shared error variance common to each individual. This allows the wave level residuals to be uncorrelated and unique to each person-wave.[11]

The multilevel model for change allows researchers to attempt to understand the average developmental course of the outcome and what variables both at the within-person and between-person levels predict that trajectory. In terms of treatment and prevention research, this allows the researcher to study not only an outcome at a specific time period, but also whether intervention alters the trajectory of substance abuse, over time. Another interesting use of this model allows for the study of intra-individual variation in outcomes. That is, the model allows us to use the person as their own control and study within person change (see Horney et al. 1995). Still, the multilevel model for change can be incredibly useful providing flexibility to the researcher, but it is limited in that it focuses on an average trajectory (while allowing individual variation around that trajectory). Researchers might be interested in homogenous subgroups that experience distinct developmental trajectories. This can be modeled via growth mixture models. This approach adds a latent class analysis to the typical multilevel model for change, such that there are latent classes (or homogenous subsets) within the sample, where the individuals within a class experience a similar trajectory, but that trajectory is different from the trajectory experienced by the other latent classes. In this way, researchers may classify different developmental processes in substance abuse or differential treatment or prevention impacts.[12]

6.7 Conclusions

In this chapter, we have reviewed the application of longitudinal methods in substance abuse research. We distinguished between repeated cross-sectional studies and panel designs, including major examples of each. We reviewed major longitudinal studies in the epidemiological, prevention, and treatment literatures and we discussed statistical methods for the analysis of longitudinal data. In this final section, we discuss the future of longitudinal data for substance abuse researchers.

[11]For an extended discussion of this model, see Singer and Willett (2003) and Bryk and Raudenbush (1987).

[12]For a more detailed discussion of growth mixture models, see Muthén et al. (2002).

One type of longitudinal data collection that could be more widely employed in substance abuse research is called ecological momentary assessment or real-time data capture. In this method, advanced technology (such as smart phones) is utilized to capture data "in real time." This is accomplished by having the participants fill out brief surveys on a smart phone or similar device at various points throughout the day. This protocol allows for the collection of an immense amount of data on the subjects and may be useful for establishing causal order among variables that change over relatively short periods of time (e.g., hours).

An additional development that would be useful for the field is the use of multiple pretests in prevention and treatment studies. Typically studies utilize a single pretest. This is limiting, because we get a snapshot of the before but ignore the process that landed the individual at the measured outcome at the start of the program. Observing multiple pretests could allow us to track the trajectory the individual was on prior to prevention or treatment and then investigate whether that trajectory is altered after intervention.

In sum, there has been a significant amount of longitudinal research employed in the study of substance abuse, prevention, and treatment. Current statistical methods allow for the analysis of longitudinal data and the future of longitudinal research in the area of substance abuse is quite promising.

References

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data* (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-046). Beverly Hills, CA: Sage.

Anglin, M. D., Hser, Y.-I., & Grella, C. E. (1997). Drug addiction and treatment careers among clients in the drug abuse treatment outcome study (DATOS). *Psychology of Addictive Behaviors, 11*(4), 308.

Bates, M. E., & Labouvie, E. W. (1997). Adolescent risk factors and the prediction of persistent alcohol and drug use into adulthood. *Alcoholism, Clinical and Experimental Research, 21*(5), 944–950.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147.

Chassin, L., Curran, P. J., Hussong, A. M., & Colder, C. R. (1996). The relation of parent alcoholism to adolescent substance use: A longitudinal follow-up study. *Journal of Abnormal Psychology, 105*(1), 70.

Chassin, L., Flora, D. B., & King, K. M. (2004). Trajectories of alcohol and drug use and dependence from adolescence to adulthood: The effects of familial alcoholism and personality. *Journal of Abnormal Psychology, 113*(4), 483.

Chassin, L., Pitts, S. C., & Prost, J. (2002). Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: Predictors and substance abuse outcomes. *Journal of Consulting and Clinical Psychology, 70*(1), 67–78.

Chen, C. M., Dufour, M. C., & Yi, H. Y. (2004). Alcohol consumption among young adults ages 18–24 in the United States: Results from the 2001–2002 NESARC Survey. *Alcohol Research & Health, 28*(4), 269–280.

Clayton, R. R., Catarello, A. M., & Johnstone, B. M. (1996). The effectiveness of drug abuse resistance education project (Project DARE): 5-year follow-up results. *Preventive Medicine, 25,* 307–318.

Curran, P. J., Muthen, B. O., & Harford, T. C. (1998). The influence of changes in marital status on developmental trajectories of alcohol use in young adults. *Journal of Studies on Alcohol, 59*(6), 647–658.

DATOS—About DATOS. (2007, September 21). *DATOS—About DATOS*. Retrieved May 18, 2014, from http://www.datos.org/aboutdatos.html

De Bono, M., & Bargmann, C. I. (1998). Natural variation in a Neuropeptide Y Receptor Homolog modifies social behavior and food response in *C. elegans*. *Cell, 94*(5), 679–689.

DeGarmo, D. S., Eddy, J. M., Reid, J. B., & Fetrow, R. A. (2009). Evaluating mediators of the impact of the linking the interests of families and teachers (LIFT) multimodal preventive intervention on substance use initiation and growth across adolescence. *Prevention Science, 10*(3), 208–220.

Ellickson, P. L., & Bell, R. M. (1990). Drug prevention in junior high: A multi-site longitudinal test. *Science, 247*(16), 1299–1305.

Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review, 5*(4), 273–297.

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2006). Cannabis use and other illicit drug use: Testing the cannabis gateway hypothesis. *Addiction, 101*(4), 556–569.

Grella, C. E., & Joshi, V. (1999). Gender differences in drug treatment careers among clients in the national drug abuse treatment outcome study. *The American Journal of Drug and Alcohol Abuse, 25*(3), 385–406.

Grunburg, N. E., & Faraday, M. M. (2002). The values of animal models to examine the gateway hypothesis. In D. B. Kandel (Ed.), *Stages and pathways of drug involvement: Examining the gateway hypothesis*. Cambridge, UK: Cambridge University Press.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112*(1), 64.

Horney, J., Osgood, D. W., & Marshall, I. H. (1995). Criminal careers in the short-term: Intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review,* 655–673.

Hser, Y. I., Anglin, M. D., Grella, C., Longshore, D., & Prendergast, M. L. (1997). Drug treatment careers a conceptual framework and existing research findings. *Journal of Substance Abuse Treatment, 14*(6), 543–558.

Hser, Y.-I., Longshore, D., & Anglin, M. D. (2007). The life course perspective on drug use a conceptual framework for understanding drug use trajectories. *Evaluation Review, 31*(6), 515–547.

Hubbard, R. L., Craddock, S. G., & Anderson, J. (2003). Overview of 5-year follow-up outcomes in the drug abuse treatment outcome studies (DATOS). *Journal of Substance Abuse Treatment, 25*(3), 125–134.

Hubbard, R. L., Craddock, S. G., & Flynn, P. M. (1997). Overview of 1-year follow-up outcomes in the drug abuse treatment outcome study (DATOS). *Psychology of Addictive Behaviors, 11*(4), 261–278.

Hussong, A. M., Hicks, R. E., Levy, S. A., & Curran, P. J. (2001). Specifying the relations between affect and heavy alcohol use among young adults. *Journal of Abnormal Psychology, 110*(3), 449–461.

Kandel, D. (1975). Stages in adolescent involvement in drug use. *Science, 190*(4217), 912–914.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95,* S5–S28.

Kellam, S. G., Mackenzie, A. C., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The good behavior game and the future of prevention and treatment. *Addiction Science Clinical Practice, 6*, 73–84.

Kellam, S. G., Wang, W., Mackenzie, A. C., Brown, C. H., Ompad, D. C., Or, F., … & Windham, A. (2014). The impact of the Good Behavior Game, a universal classroom-based prevention intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science, 15,*(1), 6–18.

Levin, E. D., Rezvani, A. H., Montoya, D., Rose, J. E., & Swartzwelder, H. S. (2003). Adolescent-onset nicotine self-administration modeled in female rats. *Psychopharmacology (Berl), 169*(2), 141–149.

Lopez-Quintero, C., Cobos, J. P. D. L., Hasin, D. S., Okuda, M., Wang, S., Grant, B. F., et al. (2011). Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: Results of the national epidemiologic survey on alcohol and related conditions (NESARC). *Drug and Alcohol Dependence, 115*(1), 120–130.

Lynam, D. R., Milich, R., Zimmerman, R., Novak, S. P., Logan, T. K., Martin, C., et al. (1999). Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology, 67*(4), 590–593.

Maggs, J. L., & Schulenberg, J. E. (2004). Trajectories of alcohol use during the transition to adulthood. *Alcohol Research, 28*(4), 195.

McKay, J. R., & Weiss, R. V. (2001). A review of temporal effects and outcome predictors in substance abuse treatment studies with long-term follow-ups preliminary results and methodological issues. *Evaluation Review, 25*(2), 113–161.

Monitoring the Future. www.monitoringthefuture.org

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*(4), 459–475.

Pentz, M. A., MacKinnon, D. P., Dwyer, J. H., Wang, E. Y. I., Hansen, W. B., Flay, B. R., et al. (1989). Longitudinal effects of the midwestern prevention project on regular and experimental smoking in adolescents. *Preventive Medicine, 18,* 304–321.

Poduska, J. M., Kellam, S. G., Wang, W., Brown, C. H., Ialongo, N. S., & Toyinbo, P. (2008). Impact of the good behavior game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence, 95,* S29–S44.

Simpson, D. D., Joe, G. W., & Broome, K. M. (2002). A national 5-year follow-up of treatment outcomes for cocaine dependence. *Archives of General Psychiatry, 59*(6), 538–544.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sloboda, Z., Stephens, R. C., Stephens, P. C., Grey, S. F., Teasdale, B., Hawthorne, R. D., et al. (2009). The adolescent substance abuse prevention study: A randomized field trial of a universal substance abuse prevention program. *Drug and Alcohol Dependence, 102*(1), 1–10.

Substance Abuse and Mental Health Services Administration. (2013). *Results from the 2012 national survey on drug use and health: Summary of national findings*, NSDUH Series H-46, HHS Publication No. (SMA) 13-4795. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Terry-McElrath, Y. M., O'Malley, P. M., & Johnston, L. D. (2013). Simultaneous alcohol and marijuana use among US high school seniors from 1976 to 2011: Trends, reasons, and situations. *Drug and Alcohol Dependence, 133*(1), 71–79.

Vesga-Lopez, O., Schneier, F. R., & Wang, S. (2008). Gender differences in generalized anxiety disorders: Results from the national epidemiologic survey on alcohol and related conditions (NESARC). *Journal of Clinical Psychiatry, 69*(10), 1606–1616.

Part III

Qualitative and Mixed-Method Approaches

Qualitative Methods in Substance Abuse Research

7

Paul Draus

7.1 Introduction

Qualitative depictions of substance use behavior and experience are nothing new. The paintings of Bruegel, which date from the sixteenth century, present rich portrayals of European peasant life, including boisterous festivals likely fueled not only by alcohol but by psychoactive poppy seeds and mushrooms baked into breads and distilled into tinctures (Palmer 2000). In ancient Greece psychotropic botanicals were utilized to inspire creativity (Hillman 2008), and Homer's accounts of the lotus eaters may be seen as a kind of ethnographic encounter involving both cultural others and the experience of mind-altering substances. In the English-speaking world, accounts of substance use experience go back at least as far as the 1670s, when a merchant seaman named Thomas Bowrey described the results of his crew imbibing a cannabis concoction called bhang along the coast of Bengal (Davenport-Hines 2001). As Page and Singer (2010) have shown, the origins of substance use research are inseparable from the sociopolitical contexts that produced them—often encounters between would-be colonizers and those they wished to understand and eventually conquer.

Perhaps the most famous and influential example of an experiential account of substance use is Thomas De Quincey's *Confessions of an English Opium Eater* from 1886 (De Quincey 2003). De Quincey covers every aspect of drug use experience in great detail, from his first exposure to opium (as a treatment for a painful toothache) to the euphoric effects it produced and the excruciating pangs of withdrawal when he tried to reduce his use. He employs what we would today call a blend of positivist and phenomenological paradigms, capturing both objective data concerning quantities of drug and mode of administration as well as the elaborate contents of his opium-inspired dreams. Even in this quintessentially individualized format, one cannot help but notice the traces of the global context; images of South Asia abound in the narrative, with all the accompanying Orientalist connotations of the Victorian era.

The explosive growth of industrial cities in the late nineteenth century contributed to the rise of the academic profession of sociology, as scholars and intellectuals sought to understand the implications of these large-scale changes in work and settlement patterns. The seismic shifts in population that accompanied urbanization and industrialization naturally brought other changes as well, including new patterns of substance use. De Quincey's experience is inseparable from London's position at the center of an expanding empire, which brought new substances such as opium within the reach of Englishmen. In the

P. Draus (✉)
Department of Behavioral Sciences, The University of Michigan-Dearborn, 4901 Evergreen Rd., Dearborn, MI 48128, USA
e-mail: draus@umich.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_7

United States, immigrants brought their habits with them, including the various European cultures of alcohol consumption as well as other drugs such as opium and marijuana. At the same time, medical science also contributed to substance use patterns through the development of products such as cocaine and heroin in the late nineteenth and twentieth centuries. With the consolidation of the medical profession and the criminalization of substance use in the early twentieth century, the peculiar combination of medical, legal, and moral discourses that now defines the policy regime of Western countries fully emerged.

While this chapter is not primarily concerned with the history of U.S. drug policy, contemporary qualitative methods cannot be understood apart from the context that shaped them. Just to give a quick example, a major issue in qualitative substance abuse research is the recruitment of participants. In a society where substance use and abuse is often illegal, stigmatized (or both) the recruitment process is inevitably complicated. Furthermore, the content of interviews themselves may be viewed skeptically, either because the reliability of participants or their willingness to speak truthfully is held in question. Thus we have the distinctive problem of "hidden" or "hard to reach" populations (Watters and Biernacki 1989; Singer 2013), people engaged in behaviors that are publically prohibited or potentially shaming.

These populations are not literally hidden, of course, as they are all around us in contemporary society. However, their status as drug users is not visibly identifiable or openly revealed. This can make them difficult to locate, count, or engage. Researchers have been quite inventive and have proven repeatedly that substance users are not only willing, but under the right circumstances, will talk quite openly about their use of alcohol and/or drugs, providing invaluable sources of information concerning substance use experiences, practices, and belief systems. Nonetheless, recruitment and reliability remain central issues in qualitative research on substance use; and policy awareness is crucial to the research process.

In the pages that follow, this chapter will first provide some philosophical background on qualitative research methods, distinguishing between three paradigmatic approaches: positivist, phenomenological, and pragmatist. It will be argued that these are better conceptualized as streams that blend and blur into each other, rather than as completely separate schools of thought. All qualitative approaches have certain things in common, namely the goal of understanding human actions from the perspective of those who engage in them, sometimes referred to as "the insider's perspective" or the "native's point of view." Nonetheless, the distinctions are important as they may contribute to different methodological approaches and distinctive interpretations of research findings. In other words, "the native's point of view" may mean very different things to different sets of people—including the "natives" themselves.

Following this brief and limited philosophical discussion, a scan of the evolution of qualitative methods for studying substance use, from classical anthropology and sociology and continuing through the postmodernism of the late-twentieth century, finally touching on posthumanist approaches informed by Science and Technology Studies (STS) and Actor Network Theory (ANT) that have emerged in the early twenty-first century will be undertaken. This section is not intended to be comprehensive, but rather to provide a theoretical framework for a discussion of specific methodologies.[1]

Specifically, this chapter will provide an in-depth discussion of the two foundational qualitative methodologies, participant observation and ethnographic interviewing, as well as the practical and ethical issues attendant to all substance use research in the context of prohibition and social stigma. Following this, it explore uses of complementary methods such as focus groups, content analysis, cognitive and ethnographic mapping, space-time diaries and geographic information systems (GIS), autoethnography, Photovoice, and community-based approaches such as Participatory Action Research (PAR).

[1]For a concise summary of the contributions of qualitative research to the study of substance abuse, including treatment and prevention, see Nichter et al. 2004.

Finally, the chapter concludes with a discussion of two proverbial elephants. One is the phenomenon of substance use itself, which will be described differently by individuals who are only touching one part of the creature's body (like the blind old men in the Jain parable). It must be accepted that all accounts of substance use are themselves partial, and that each qualitative description has something to contribute to our understanding of the whole elephant, even if no two are exactly alike. The other proverbial elephant is one that has already been referenced—that of the current drug policy, sometimes described simply as the "Drug War", which is largely defined by legal prohibition and punitive criminalization.

7.2 Philosophical Background: The Three P's (Plus a Couple of Posts)

While a discussion of qualitative research paradigms could have many starting points, I choose to begin with the sociologist Alfred Schutz. This may seem an odd entry point, as Schutz was not himself an ethnographer. However, Schutz's distinctive contribution was to build a philosophical foundation for research in the social sciences. This contribution is based on the idea that the social world was not merely an object that could be studied in the same manner as physical nature, but was constructed intersubjectively—by and through the actions and beliefs of people existing in relationship to each other (Dreher 2011).

The English translation of Schutz's *The Phenomenology of the Social World* was published in 1967, around the same time as Berger and Luckmann's *The Social Construction of Reality* (1966) and Garfinkel's *Studies in Ethnomethodology* (1967). These works and numerous others ushered in a new era of interest in the theory and methodology of qualitative research. Drug-related research also experienced a significant resurgence at this time. In particular, the labeling theory of deviance as advanced by sociologists such as Becker (1964) and Goffman (1963) disrupted the notion that drug users were morally and psychologically distinct from "normal" members of society—that they were merely "mad", "bad", or "sad." Instead, they focused on the marking of difference and the reactions to that marking, both by "normals" (Goffman 1963) and "outsiders" (Becker 1964). Becker's (1953) article "Becoming a Marihuana User" challenged the idea that marijuana users were defined by inherent predisposition, arguing instead that both the behavior and the accompanying social role were learned through a complex series of social interactions. Furthermore, Becker maintained that smokers had to learn how to recognize the drug's effects and to know that they were "high."

Even from the standpoint of the twentieth century, one can see how this perspective unsettles assumptions about drug use and drug users. It suggests that the characteristics of either the pharmaceutical agent or the person ingesting it may be less important than the context that surrounds them, which shapes their interpretations and thus their actions and behaviors. The period from the 1960s to the 1980s saw a proliferation of qualitative approaches to understanding drug subcultures and drug-using careers. These studies offered windows into the daily lives of drug users across a wide range of settings, and highlighted the importance of social context for understanding problem drug use. Zinberg's (1984) tripartite formulation of "drug, set and setting," based on research conducted with soldiers who had become physically addicted to heroin while serving in Vietnam but were able to abandon use completely upon returning to the United States, further advanced the idea that the social-environmental "setting" was the most powerful factor influencing actual patterns of use.

Phenomenological research is concerned with conveying the essence of subjective experience, as De Quincey sought to do in his accounts of his own opium-using experience (De Quincey 2003). Not all qualitative methods are necessarily phenomenological in this sense. Social research is sometimes divided into "emic" and "etic" approaches. Emic approaches seek to understand and convey the perspectives of individuals and

their life-worlds. This is a concept from Schutz (1967) which encompasses not only a person's daily life, but their past experiences, relationships, and subjective imagination. Etic approaches, on the other hand, are concerned with objectively describing the world as it is, using terms that are generalizable. Though the emic/etic distinction is sometimes equated with the qualitative/quantitative divide, this is overly simplistic; a false binary.

In fact, many qualitative researchers do employ a positivist paradigm, even if they use emic approaches. For these researchers, qualitative methods may simply provide another means to the same end, which is accurate information about social belief and practices. Anthropologists who interview informants concerning their kinship networks, for example, are interested in gathering data that is objectively correct. However, in cultures without written records, interviewing individuals may be the only way to obtain such data. The same is true in drug subcultures, as many trends in drug use change and evolve very quickly and are not officially recorded anywhere. Therefore, one needs to "go to the source" to obtain accurate information about who is doing what and where and how they are doing it. While the use of qualitative methods in epidemiological research is invaluable, the positivist orientation in that field requires that such data be legitimated by other sources if possible. Thus, we have an emphasis on triangulation, the gathering of information from multiple sources in order to strengthen the reliability of one's findings.

While positivism has had a powerful influence in all of the social sciences, other paradigms have also informed the development of qualitative methods. The Chicago School sociologists of the early and mid-twentieth century, for example, were influenced by the philosophy of pragmatism, especially the works of Mead, Dewey, and Blumer (Snell 2010). This viewpoint is succinctly expressed in the so-called Thomas theorem: "If men define situations as real, they are real in their consequences" (Thomas and Thomas 1928). While this is sometimes cited as an example of the interpretive framework, and therefore contrasted with positivism, the emphasis of the statement is as much on the consequences as the interpretation. This reflects the problem-solving orientation of much of the Chicago School's work.

It is worth noting that Thomas's observation was contained within a study of problem behaviors among children, much of which was devoted to subjects such as the treatment of maladjustment and delinquency in schools. The Chicago School also produced the first systematic study of substance use, Bingham Dai's *Opium Addiction in Chicago* (1937), which extensively employed qualitative interviews in conjunction with secondary data analysis to describe the social contexts and geographic patterns of opium use in the early twentieth century city. Like Thomas and Thomas, Dai was interested in using research to address a pressing social problem.

For this purpose, it was not necessary to understand the whole experience, but simply to gain a working knowledge of the key terms and frameworks of understanding employed by group members. This provided the basis for communication and also contributed to more effective programs and policies directed at specific groups. This pragmatic orientation persists in much of substance use research, which is ideally used to design interventions to either prevent or address the health consequences of substance misuse. In other words, it is applied research intended not to provide perfect knowledge but to guide practice and policy. From this standpoint, a central purpose of qualitative research is to obtain functional knowledge of another culture or distinctive subgroup—not just for its own sake, but in order to do something about it.

Researchers in this era tended to view drug use as a subcategory of the sociology of deviance or to portray drug users as separate and distinct cultural groups. Whereas the phenomenological, positivist and pragmatist orientations had their differences concerning the proper goals of social research, and the means to achieve those goals, they tended to share a belief that research was worth doing and that it could be accomplished by

those who were trained properly. In classical anthropology, for example, culture was an objective reality that could be systematically investigated, analyzed, and understood. At the end of the process a text would be produced that represented the outcome of these methods, which could then be judged on its own merits and used to inform policy.

This consensus was disrupted by the emergence of a fourth "P", that of postmodernism, in the 1980s and 1990s. Postmodernism is not itself a unified paradigm, but an umbrella term for a wide range of critical perspectives. For the purposes of this discussion, postmodernism is significant because it entailed a fragmentation of consensus concerning the appropriate objects of social research, its methodologies, and the nature of the social world itself. In the postmodern view, culture is a much more fluid construct, and both the interests of researchers and the role of the text are called into question.

According to Smith (1992), "The postmodern ethnographer views the researcher as a kind of translator, broker, or mediator between the world of social practice and the way the practices of everyday life are interpreted in the world" (p. 508). One of the most influential works of postmodernism, relative to ethnography, is Clifford and Marcus's (1986) *Writing Culture*, which contends that writing is itself an act of cultural construction that shapes the very social world it intends to describe. In this view, there is no such thing as objective observation, but only various texts vying for dominance in the field of representation.

It is no accident that the postmodernist movement in academia coincided with identity-based social movements and postcolonial political movements: which embraced a common critique of institutionalized authority and its representations of "the other." Though criticisms of anthropology and sociology focused on the role these fields had played in producing (or reproducing) narrative subjugation of women, ethnic minorities, and other disempowered groups, the implications for the study of substance use and abuse were also significant. In the wake of postmodernism, it became much more difficult to maintain the position of dispassionate researcher, without accounting for one's power, position, and interest in the formulation of research questions, one's chosen methodology, and the analysis of the results.

To make things even more complicated, another "post" has emerged in recent years: that of posthumanism. Posthumanism emerged from science and technology studies, specifically the work of Bruno Latour, John Law and others, who have developed an approach called Actor Network Theory, or ANT (Latour 2005; Law 2009). What ANT offers, in relation to the study of substance use behavior, is a way of talking about the role of substances and environments that escapes from the circular traps of the structure and agency debate. According to Duff (2011), ANT offers promising directions in terms of qualitative methodology, specifically in the analysis of drug use contexts.

7.3 Substance Use and Mixed-Methods Research

In spite of the "post" critiques, and the recent renaissance in qualitative methods (Gobo 2005) notwithstanding, much of social science is still dominated by a straightforward positivist approach. In drug research, qualitative methods are often employed primarily in a complementary fashion, providing interpretive depth and specific detail to cross-sectional quantitative studies (Moore 1993). Often qualitative methodology becomes the vehicle through which postmodern perspectives may be expressed. Some drug researchers maintain that the variety of drug subcultures and their incessant evolution necessitates a more flexible approach that aligns well with postmodernism's view of the fragmentation of social reality (Golub et al. 2005). Others have argued that historical circumstances have spurred the adoption of mixed-methods approaches, as in the case of the HIV/AIDS epidemic.[2]

[2]A thorough discussion of mixed-method approaches to the study of substance abuse is presented in Chap. 10 of this volume.

7.4 Drugs, AIDS, and Ethnography

Responding to the deadly syndemic of intravenous drug use, sexual behavior, and infectious disease, in the context of growing social inequality and urban disinvestment, researchers developed a wide repertoire of techniques for capturing the dynamics of drug use behavior and associated health risks. Baer et al. (2003) referred to this as the "postmodern" period in illicit drug research, in contrast with the "modern" period associated with the Chicago school and classic cultural anthropology. Page and Singer (2010) provide an even more detailed discussion of these phases in the development of drug ethnography.

For the purposes of this chapter, the HIV/AIDS epidemic was extremely significant because it forced researchers to seriously consider the interconnection between "hidden populations" of drug users and everyone else (Trotter et al. 1995). Beginning in the late 1980s, the National Institute on Drug Abuse (NIDA) through the National AIDS Demonstration Research (NADR) project, provided increased funding for innovative research while also emphasizing the practical application of research findings. The use of qualitative methods to explore dimensions of risk and inform policy (Carlson et al. 1995), sometimes called an "ethno-epidemiological approach" (Clatts et al. 2002), emphasized the pragmatic application of qualitative methods to identify small-scale trends and evolving risks related to drug administration (Clatts et al. 1999, 2007).

The late twenty and early twenty-first centuries thus saw a proliferation of intertwined qualitative methodologies including focus groups, social network analysis, and ethnographic mapping—often adapted to address particular emergent drug trends within specific subpopulations. Sampling methods used for qualitative research may also vary widely, though snowball sampling (Biernacki and Waldorf 1981) and targeted sampling (Watters and Biernacki 1989) are probably the most common. Respondent-driven sampling has been increasingly employed in recent years because of its success in accessing hard-to-reach populations (Heckathorn 1997; Draus et al. 2005; McKnight et al. 2006; Daniulatyte et al. 2012). Some of these methods are discussed below, accompanied by relevant examples. First, however, it is necessary to review the two foundational qualitative methods: participant observation and ethnographic interviews.

7.5 Participant Observation and Substance Use

According to Schutz, "all genuine understanding of the other person must start out from Acts of explication performed by the observer on his own lived experience …" (1967, p. 113). Schutz describes how one might attempt to understand the mind of the woodcutter by putting ourselves in his shoes and mimicking his movements. The social scientist seeks to make sense of the "common sense" understandings of the social actor, in part through his/her own experience. But this is only a starting point; the investigator must also seek to understand something about this woodcutter, where he comes from, and what leads him to cut the wood. For this purpose, it is imperative to talk to the woodcutter as well. Direct observation of social actions and direct access to those common sense interpretations are both necessary for the social scientist to analyze the social world in a systematic way. The two primary methods employed by qualitative researchers for this purpose are participant observation and ethnographic (or qualitative) interviewing.

Participant observation is both the easiest research activity to engage in and the hardest to do well. In a sense, all individuals are participant observers as they go through the course of their daily lives, and in contemporary society they are regularly exposed to contexts other than those with which they are intimately familiar. For instance, even as I write this, I am sitting in a coffee shop located south of the city of Detroit, inadvertently eavesdropping on the group of individuals seated next to me. While this is not what I came here to do, I am in fact engaged in a

form of participation observation while minding my own business, and track the currents of conversation, which range from educational and work experiences to the problems posed by the local economy. One of them is a Vietnam veteran, rides a motorcycle, and has worked on a number of Hollywood films that have been filmed locally. I already perceive that an interesting social world surrounds each of these four individuals (two men and two women), and that they are intersecting here but will soon return to their own regular domains.

A study could potentially be constructed around any of them, or around the space that emerges inside the coffee shop. Mitch Duneier's *Slim's Table* (1992) is an example of such a study. Duneier began just by sitting and watching, day after day. Only after he became familiar —seen as a "regular" customer—was he able to engage in the in-depth conversations and interviews that eventually formed the basis of his book. Returning to our discussion of philosophical foundations, ethnographic awareness emerges at the moment when we recognize ourselves as social actors constantly engaged in a complex intersubjective conversation with others. In other words, we come to see that the social world is not simply a natural fact that exists on its own, but is in fact constructed by each and every one of us on a moment-by-moment basis. At the same time, this world is marked by stability as well as fluidity, and there are static as well as dynamic factors to consider.

Participant observation enables observation of this process of social construction as it proceeds, while at the same time identifying those relatively static factors, such as coffee shops and work places, laws and languages, economies and institutions that shape and constrain interactions the way that banks constrain a river. It also raises awareness of the fact that riverbanks, like governments and financial institutions, are only solid and static until they are not—they can potentiallybecome as fluid or ephemeral as the swirl and flow of the water. Participant observers must seek to enter the stream, but not to get caught up in the flow. Rather, they must extract themselves, literally or figuratively, to stand on the banks and make sense of what they saw, heard and felt.

From a positivist framework, the focus might be on how the flow of the water is defined by the river, and with identifying the patterns of flow and their actual and potential trajectories. If the analogy is applied to drug use, it might be advantageous to simply characterize drug use careers or describe drug use contexts in order to identify factors that increase or decrease risk, either for the individual or society. On the other hand, as phenomenologists, it might be logical to seek to truly understand what the drug user is feeling, what motivates him/her, what satisfaction is gained, what suffering is relieved when he/she ingests that particular substance.

Finally, both positivist and phenomenological insights might be put to practical use in the pragmatic, applied tradition that defines much of American social science, and drug use research in particular. In a 2002 article, Bourgois commented on the difficulty of maintaining dialogue between these very different paradigms. However, Bourgois and Schonberg's (2009) *Righteous Dopefiend* exemplifies the integration of participant observation methodology with positivist, phenomenological and pragmatist orientations. Bourgois, an anthropologist, and Schonberg, a photographer, spent 10 years in the company of homeless heroin and crack cocaine users in and around the streets of San Francisco, observing and documenting daily life (and death) with the intent of highlighting the impact of individual risk behaviors as well as public policies. In the book, quotes from interviews, field notes, photographs, and the author's narrative all contribute to a well-rounded and complex portrait of their subjects' daily struggles.

Though every process is different, participant observation studies often go through a series of phases. Spradley (1980) presents a cyclical model that is contrasted with the more linear model of traditional positivist social research. Spradley's Ethnographic Research Cycle begins with the selection of a social setting, the asking of ethnographic questions, the collection of data, analysis of data, and the formulation of new questions. The first phase, often referred to as the

"breaking in" phase, involves entering a new context, establishing a presence there, and becoming familiar with the routines and key players in that setting.

For drug research, this process is once again complicated by the hidden—that is, illegal or stigmatized—nature of the activity involved. This is different for alcohol and tobacco research, and it is changing rapidly for marijuana, but for illicit drugs such as heroin, cocaine, and methamphetamine it is generally understood that active users do not want to be easily identified. Gaining access to those populations may require more time, effort and expense, and even then there may be limits to the access one can get. Sometimes one may be fortunate enough to meet an individual "key informant" and gain access through them, as in the famous example of "Doc" described by Whyte (1943) in *Street Corner Society*. In other cases, a contact may be peripheral to a network but still serve as a valuable "cultural broker" (Moore 1993). Indigenous field workers, usually former drug users themselves, may also be employed by a research project for the purpose of establishing connections with active drug users.

In some cases, drug use may simply be a part of the setting, and not the subject of a study. In other cases, the researcher has a specific goal of accessing a hidden or hard to reach population. I have had both experiences in my own career. My dissertation research focused on the social context of tuberculosis on late twentieth century Chicago. I spent many hours every week in neighborhoods with high rates of tuberculosis, and my fieldwork and interviews were primarily concerned with the experience of tuberculosis, the environments that contributed to it, and its potential causes and its treatment. However, it just so happened that the neighborhoods where tuberculosis patients lived also contained very active drug markets. Many of the patients that I saw had addiction issues, either with alcohol or illicit drugs such as crack cocaine and heroin. Therefore, while substance use was not the subject of my study, it inevitably became a part of it. Furthermore, it quickly became apparent that the environmental factors shaping tuberculosis where inseparable from those contributing to problem substance use. By the time I wrote up my results, access to the population was not an issue because I was already well known to my participants.

However, when I was hired to be the Project Ethnographer for a NIDA-funded study of illicit stimulant use in rural Ohio several years later the situation was reversed. Now I was charged with implementing a Respondent-Drive Sampling (RDS) plan. I had no familiarity with the areas where the recruitment was to occur, and the atmosphere of the small Midwestern town rendered drug use, and drug users, even less visible and accessible than they were in inner city neighborhoods (Draus and Carlson 2006, 2007, 2009). For those not familiar with RDS, it is similar to other tested recruitment methods such as snowball sampling, but it requires limiting each initial research subject (called "seeds") to only three referrals. The process of recruitment was complex and involved a number of steps and the application of ethnographic methods throughout, as each local setting required a recalibration of recruitment methods and assessment of potential barriers, especially in terms of trust (Draus et al. 2005).

While this was ultimately an interview-based study, it was necessary to conduct extensive participant observation fieldwork in a series of small towns—essentially starting from scratch in each place. Engaging in impromptu conversations on front porches and in yard sales, at county fair carnivals and summer music festivals, in city parks and on the steps of courthouses, in tattoo parlors and roadside bars, we sought to learn about the social context of the town while also seeking out those who were willing to talk about issues related to drug use. Unlike many substance abuse studies, recruitment of participants from treatment centers was avoided, in part because residential treatment did not exist in most of these small towns, but also because of a desire to engage active (not in treatment) stimulant users.

This example highlights the interconnection between research and recruitment in ethnographic studies. While sampling is often treated

as a separate element of research, in fact qualitative interviews, participant observation and subject recruitment are mutually reinforcing activities in qualitative studies.

7.6 Ethnographic or Qualitative Interviews

Following participant observation, interviewing is the most basic and fundamental qualitative research methodology. In fact, it probably surpasses traditional participant observation fieldwork in terms of the number of studies that employ it, especially in substance use research. There are a couple of reasons for this. One is that interviewing is often simply more efficient and less time-consuming than long-term immersion in a fieldwork setting. In studies that are government-funded and have set timelines and benchmarks, qualitative interviews are also a more convenient way to demonstrate what was done. Though many (if not most) qualitative researchers are trained in the "pure" ethnographic method (I put it in quotes because I am not quite sure that such a thing really exists), in practice researchers must sometimes rely on shortcuts, because "true" ethnography is extremely time and labor-intensive. In other words, while it might be preferable to spend a full year simply hanging around drug users and getting to know a particular scene, the same basic information may be obtained by interviewing someone who has been in the scene for longer than that, illustrating the potential tension between the pragmatist and phenomenological approaches.

A second reason why interviewing is such a and fundamental qualitative research methodology is that interviews lend themselves better to quantification, generalizability and comparability, considerations which are important when employing a positivist paradigm or when seeking the approval of positivists, either for funding or publication purposes. Not only can interviewers themselves be counted (providing a specific "n" to cite in research), but responses can be codified and counted as well. Many reviewers, both for professional journals and grant-making agencies, are trained in positivist methods and are not interested in funding what they call "fishing expeditions." From a pure ethnographic standpoint, of course, there is nothing inherently wrong with a fishing expedition. It is possible to catch something entirely unexpected that will potentially shift perspective on a situation and offer new and original insights. That can still happen in a funded research project with a finely detailed list of Specific Aims, but it will not have gotten funded for that purpose.

Third, from a phenomenological standpoint there are things about another person's experience that cannot be understood in any other way —even in cases where the researcher actively participates in the same context, as their lived experience may be different. Therefore it is necessary to ask people directly, and to allow them to speak for themselves. Interviews provide an ideal format for this sharing of experiences. Finally, there may be behaviors that researchers are not able to engage in personally for ethical, legal, health, and/or other reasons. Substance abuse would usually fall into this latter category.

As with other qualitative methods, interviews may be informed by multiple paradigms. Shaw's classic interview-based study "*The Jack Roller*" (1930) takes the reader on a journey through one young man's experience in a descriptive, empirical sense while also conveying something of his quality of mind, with the pragmatic purpose of informing policy concerning juvenile delinquency and urban environments. Bingham Dai's (1937) groundbreaking work on opium use in Chicago, referenced above, was based primarily on interview data collected inside psychiatric hospitals, as was Lindesmith's research on opiates and addiction (1968).

The use of interviews to understand specific mindsets, while also addressing practical problems, is a characteristic of much qualitative research on drug use because of the social disorder and individual risk that is often (rightly or wrongly) associated with illicit drug use. Quintero and Estrada's (1998) study of injection drug use and machismo in a US–Mexico border city and Carlson's (1999) study of gendered associations with heroin and cocaine, referred to as

"boy" and "girl" by Ohio drug users, are two good examples of how researchers may incorporate both positivist and phenomenological approaches with a pragmatic purpose. The classic text on ethnographic interviewing is that of James Spradley (1979). Specifically, Spradley described the ethnographic interview as a particular type of a speech event, similar in many respects to a casual conversation but with a distinctly different purpose: that of eliciting the insider's perspective or point of view. Appealing to the symbolic-interactionist tradition of sociologists such as Blumer, Mead, Cooley, and Thomas, Spradley argues for the centrality of shared meanings that directly influence behavior. Recalling the pragmatist tradition, he described culture as a cognitive map that group members use to navigate daily life. A map is not determinative. Rather it implies a set of flexible guidelines: "It serves as a guide for acting and for interpreting our experience; it does not compel us to follow a particular course" (Spradley 1979).

However, because these maps are not written down, but are carried around in people's head, the ethnographic interview is one of the most effective means of elucidating their contents. To do this, Spradley lays out a whole sequence of techniques, from the "grand tour" question, which invites participants to lay out the broad parameters of a particular place or social experience, to the more directed "example" or "experience" questions which probe for specific details within that broader canvas (Spradley 1970). While many others have refined Spradley's method, and every researcher must adapt it to her or his own circumstances, it remains essential as a practical outline for conducting ethnographic interviews. Bevan (2014) for example, has written on the practice of phenomenological interviewing, building off the framework developed by Spradley. Furthermore, Spradley provides an ethical rationale for the ethnographer's work, arguing that researchers have two central purposes: understanding the human species, and serving the needs of humankind (Spradley 1979).

Spradley's work is particularly significant for the purposes of this chapter because he made his reputation through an in-depth ethnographic study of itinerant alcoholics in Seattle. In a preface to the 30-year anniversary reissue of *You Owe Yourself a Drunk* (Spradley 1970 [2000]), Singer credited Spradley with bringing applied anthropology on substance use to the fore of the field at the time of its publication, while also enduring the critiques of postmodernism in the intervening years. The wealth of research on substance use conducted since 1970 by anthropologists, sociologists, and others employing participant observation and qualitative interviews is a testament to the enduring relevance of the approach that Spradley helped to define.

In my own work, the model provided by Spradley has been extremely useful, both in terms of methodological structure and ethical purpose. From my years as a graduate student engaged in long-term daily interaction with tuberculosis patients in Chicago, my subsequent work as a full-time ethnographer of substance use in rural Ohio, and my current role as a faculty member engaged in a variety of community-based research projects, I have relied on the form of the ethnographic interview as a central tool in investigating the life-worlds of others.[3] Spradley's injunction that research ought to at least attempt to shape the world in a way that is beneficial for other human beings remains a constant challenge, as it should be for those of us who conduct research among society's most marginalized members.

While participant observation and ethnographic interviews continue to form the spine and muscle of most qualitative research projects, a number of other techniques have emerged in recent years. Each of these builds on the same philosophical framework as the methods discussed above, but also offers additional avenues for understanding the perspectives of particular

[3]Recent work includes qualitative interviews with former street sex workers in Detroit, Michigan (Draus et al. 2015a, b). In an analysis of women's accounts of their transition from active substance use to a recovery-oriented existence, for instance, the importance of social networks and daily routines were identified as key components of the recovery process, as women reassembled their identities to align with the goals of recovery.

groups. Four of these will be discussed briefly—content analysis, focus groups, social network analysis, and visual ethnography, with some elaboration on their potential contributions to substance use research.

7.7 Content Analysis

Content analysis has a long history in qualitative research. One of the most influential early examples is Thomas and Znaniecki's *The Polish Peasant in Europe and America* (1958[1918–1920]), which relied on the systematic analysis of letters exchanged between Polish immigrants and their friends and relatives at home. Much content analysis related to substance user research has focused on depictions of drugs in the media or in legal or clinical documents (Primack et al. 2008). In some cases, content analysis has been performed on qualitative interviews themselves, using computer software to quantify references to specific themes such as gender and intimate relationships (Henderson et al. 1994).

7.8 Focus Groups

Focus groups have their origins in social psychology and sociology in the early twentieth century but through much of their history have primarily been employed for marketing and political purposes. In these settings, focus groups were intended to elicit the viewpoints of target demographics on specific products or issues. A familiar use of focus groups includes those assembled following presidential debates to understand how, for example, "undecided voters" are reacting to the statements or performances of different candidates. While not a new methodology in the social sciences, focus groups began to gain fresh currency in the 1990s, with expanded use in areas such as public health and AIDS research (Kitzinger 1994). What distinguishes focus groups from individual interviews (aside from the fact that multiple participants are involved) is the purpose of elucidating a specific issue or problem from a group perspective. The specific issue or problem is what gives the group its "focus", but the "group" deserves just as much emphasis, because what emerges in the focus group context may be something quite different from what comes out of a series of single interviews. In previous studies, for example, I have used focus groups to investigate trends in substance use or to illuminate community-based barriers to subject recruitment (Draus et al. 2005). Likewise, Valdez and Kaplan (1998) used focus groups to address selection bias in recruitment of Mexican-American gang members. In each of these cases, researchers are concerned with group-level phenomena: collective beliefs and experiences related to those particular issues.

The dynamic of the group can produce something greater than the sum of its parts, as the participants interact and converse and elaborate each other's thinking or add to the limited base of a single experience. On the other hand, some group members can dominate others, sway them to agree with a particular point of view, or else silence them completely. The task of the facilitator, therefore, is just as crucial as that of the qualitative interviewer, but may require some different skills. According to Agar (2006), focus groups are not themselves sufficient to constitute ethnography because they are narrow gauge rather than holistic. On the other hand, focus groups can fit very well into a more targeted, problem-focused or pragmatic approach, such as those studies aimed at understanding needle-sharing behaviors among localized groups of intravenous drug users. They can also generate a lot of interesting ideas in a hurry, which undoubtedly accounts for some of their appeal. In any case, the use of focus groups is probably best approached as a complement to other qualitative methods.

7.9 Social Network Analysis

Although social networks had always been implicit in studies of drug use, the systematic integration of qualitative methods with social network analysis has been a relatively recent development. Again, the urgency of the

HIV/AIDS epidemic catalyzed innovation in this area, contributing to a NIDA research monograph that was issued in 1995 (Needle et al. 1995). While social networks analysis is not inherently qualitative (in fact much of social networks research is heavily mathematical), it can be integrated relatively smoothly into other qualitative interviews. In addition, the combination of a social networks generator with an ethnographic interview provides significant details on everyday life while also generating network data that can be visualized graphically or analyzed mathematically. A social networks generator is basically a list of social contacts, which can be more or less detailed depending on the time constraints and purpose of the interview or study. Because so much of drug-using behavior is related to one's social networks, understanding those networks as structures and factors in and of themselves also offers another potential route of intervening in dangerous drug use or related high-risk behaviors (Neaigus et al. 1994; Weeks et al. 2002; Valdez et al. 2008).[4]

In my Detroit-based research, I have utilized a social networks generator alongside a daily routines interview, which builds off Spradley's "grand tour" framework to capture nuances of participants' everyday lives while also establishing mobility patterns and relationships that can be mapped (Draus et al. 2010a, b, 2012). The recollection of people, places, and things encountered in the course of one's daily round effectively function as prompts similar to Spradley's "example" and "experience" questions. Likewise, in a study of former street sex workers, I used these methods to develop indices of recovery from substance use, based on the hypothesis that social integration from the margins to the mainstream would be accompanied by documentable changes in social networks and mobility patterns (Draus et al. 2015a). This is another example of how a qualitative study may have positivist purposes, without losing its phenomenological core. While the structure of social networks may be described using objective or etic terminology, it may be argued that the content of those networks cannot be understood without the insider's perspective.

Another expanding area of social networks analysis in substance use research has to do with social media. Social media such as Facebook©, Twitter©, and others may provide rich sources of qualitative data (Scott 2013). Using the techniques of content analysis as well as those of social networks analysis, researchers are just beginning to explore the world of meanings associated with the Internet. Nicholson et al. (1998) conducted early exploratory research on the World Wide Web as a means of accessing hidden populations and found that there was a bias towards white, better educated and male participants. However, in the succeeding years the use of the Internet has expanded exponentially, as have social media. This promises to be an expanding site of qualitative research in years to come.

7.10 Visual Ethnography

The last area is that of visual ethnography, which also includes a wide variety of methodologies, from participatory approaches such as Photovoice (Wang and Burris 1997; Wang 1999), to more traditional documentary photography or filmmaking. I employed photography in a participatory manner in my early research on tuberculosis (Draus 2000, 2004), primarily as a means for individuals to document their environment as a vehicle of risk and a repository of relationships. As mentioned above, Bourgois and Schonberg (2009) employed photography alongside participant observation and in-depth interviews to investigate the context of homeless men in San Francisco as well as to interrogate policies and their effects. Of course, photography has its own ethnographic tradition, a powerful example being Larry Clark's *Tulsa* (1971), which depicted the lives of young people who were injecting amphetamine and with whom Clark himself had a long history.

[4]By definition, social network analysis is a transdisciplinary endeavor (Valente et al. 2004). See chapter by Reingle and Akers (this volume) for additional information on the importance of a transdisciplinary framework for substance abuse research.

As with the Internet and social media, access to handheld technology has greatly expanded and the potential for people to document and describe their own environments visually offer enormous possibilities for research, and we have recently utilized Photovoice in Detroit as a means for young people to depict the aspects of their environment that contribute to feelings of risk or danger as well their sense of safety and pride (Draus and Roddy 2013). In an ongoing study examining the changing patterns of marijuana use in the wake of medical marijuana legislation, we are employing photography as an ethnographic tool to understand evolving social contexts in the Detroit metropolitan area, especially as they relate to geography and racial segregation (Roddy and Draus 2014). Furthermore, we are exploring the potential for smart phones to be utilized as participatory tools of ethnographic research on the changing landscape in Detroit as it relates to patterns of substance use and drug-selling markets. This technology also allows the simultaneous collection of social network, visual, and geographic data.[5]

7.11 Qualitative Data Analysis Software

In this new era of research, traditional methods of participant observation and ethnographic interviewing will be greatly enhanced by technology. Likewise, software designed to assist in the analysis of qualitative data continues to evolve. Software packages such as Atlas TI, NVivo, HyperResearch, and others offer a wide range of features, including the capacity to consume a variety of forms of data, from textual documents to audio and video-recorded interviews or focus groups. In addition to traditional coding functions, these applications can perform complex queries as well as produce theoretical diagrams, word clouds, and so on.

Some of these tools can be used to convert qualitative source material into quantified outputs. Nonetheless, the soul of qualitative research will remain the same: the accurate and respectful representation of the perspective of the social participant, whether that be the insider, outside, or persons or groups in between. Likewise, the tensions and conflicts inherent in qualitative research on substance use will persist, and must be negotiated over and over again. It is to these matters that I turn in the next and final section.

[5]See also Chap. 9 in this volume for information of the use of GIS in substance abuse research.

7.12 Conclusion: The Proverbial Elephants

The traditional Jain tale of the learned blind men and the elephant is widely used in the social sciences, and for good reason. One of the men feels the elephant's side, and pronounces it to be like a wall. Another feels the tusk, and says it is more like a spear, while the one who touches the trunk compares it to snake, and so on. The conclusion of the story, told in the form of a poem by John Godfrey Saxe (1816–1887), goes like this

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

This proverb is reflective of the complex nature of social reality itself. For every set of observable facts, there may be numerous interpretations, each of which is valid within its own set of framing assumptions. We have already discussed the different paradigms that may inform qualitative research methods, and it is readily apparent that these can lead reasonable people to disagree on what any set of research findings actually means. Furthermore, as the parable also reveals, the facts themselves may vary depending on the angle of approach–whether one is touching the ear, the leg or the trunk.

The second proverbial elephant in the room, when discussing drug research, is that of official state policy. The most influential theorist of addiction in the mid-twentieth century was a qualitative researcher: Alfred Lindesmith, who argued for the consideration of subjective psycho

logical states and social contexts as fundamental components of drug-using behavior. Lindesmith, who was trained by sociologist Herbert Blumer at the University of Chicago, employed in-depth interviews with long-term opiate users to develop his theory, which was informed by Blumer's symbolic-interactionist approach. Lindesmith was also influenced by the pragmatist tradition in Chicago, and was a sharp critic of punitive policies towards drug users which he viewed as profoundly counterproductive. In fact, he was individually targeted by Harry Anslinger and the Federal Bureau of Narcotics (Galliher et al. 1998). The conflict between Lindesmith's sociologically informed position and that of the US government was to prove emblematic of the relationship between research and state policy into the twenty-first century.

These tensions have not disappeared in the contemporary world of ethnography or substance use research. The issue of marijuana provides one salient example. Social policy and attitudes around marijuana are evolving very quickly. Its use is quite prevalent across different social demographics and many of its users regard it not only as harmless, compared to other psychoactive drugs, but as medically beneficial (Parker et al. 1998, 2002). Federal lawmakers and agencies, on the other hand, are still dominated by a drug war paradigm, seeing the expanded use of marijuana as not only harmful in its own right, but potentially leading to other more dangerous forms of drug use. Ethnographers who seek to honestly convey the perspectives of their research subjects may find themselves in tricky situations relative to both funding sources and community-based partners who see marijuana as part and parcel of a broader drug threat. I had a glimpse of this at a recent meeting of public health, law enforcement and treatment providers in Detroit, a group that was originally organized to address overdose risks associated with the powerful pain reliever called fentanyl. In this particular meeting, the issue of medical marijuana was portrayed in different ways by the various people at the table. One woman, for example, described it as part of a "leaky bucket" strategy designed to pave the way for broader drug legalization. Others focused on the discrepancies between enforcement strategies at the local, state and federal level, while still others saw it primarily in terms of harm reduction.

Researchers and social scientists who do community-based work need to be cautious when entering into this fractious terrain. Ethnographic researchers have sometimes violated closely held beliefs of the medical and legal establishments by asserting the relative normality of substance users and the potential for environments and policies to reduce harm rather than aggravate it. In some ways research on marijuana is more complicated than research on drugs such as heroin or practices such as intravenous injection, which everyone (including users) generally agrees are fairly risky endeavors. Marijuana use, on the other hand, falls somewhere between the zone of normalization and stigma, between risky behavior and recreational activity. In this sense, it resides at an opposite extreme from research on substance use related to the transmission of HIV which tended to unite researchers, practitioners and policymakers due to the urgency of understanding and reducing risk. Harm reduction arguments stem from an analysis of drug use that sees punishment and repression as leading to more risk, rather than less, but still there is an underlying assumption that drug use is risky.

For example, one might seek to examine whether increased marijuana use due to legalization or decriminalization is associated with risky sexual behavior or increased risk of using more dangerous drugs. Some studies show that medical marijuana laws may lead to increased drug potency and contribute to negative health outcomes (Sevigny et al. 2014). Our early findings on patterns of marijuana use and procurement in the wake of medical marijuana legislation in Michigan (Draus and Roddy, unpublished manuscript) seem to indicate that both race and geography are influencing how people use and buy. White users, at least at the time of this writing, are more likely to take up the medical language and to abide by the new law in order to be protected in their use.

African American users, on the other hand, are less likely to define their own marijuana use

as "medical", to purchase marijuana that is described as a medical product, or to pursue a state-issued medical card. This may mean that they are at greater risk of prosecution for marijuana possession. On the other hand, white users who do pursue medical status, as either users or growers, may not be immune from arrest or prosecution. In fact, they may be at elevated risk, especially in some jurisdictions, because they have become more public in their use and therefore easier to raid. These sources of risk may be entirely separate from the question of health risk resulting from marijuana use itself. On this question, black and white users are fairly unified in the sense that neither group assigns much (if any) risk to marijuana use.

Researching this complex issue using a qualitative paradigm requires an understanding of the viewpoints held by all these various participants, and to point out the gaps between popular, professional, and policymakers' perspectives. At the same time, researchers who employ a positivist or pragmatist approach will seek to filter through the welter of viewpoints and experiences to arrive at a set of reliable and actionable facts from which policy recommendations can be made. This may result in a research process that sometime resembles, in the memorable words of Burawoy (2014), a contact sport, operating at the intersecting edge of the academic, political, and practice fields. However, as with the multiplicity of paradigms and perspectives that bump up against each other in qualitative research, it may be argued that this jostling and argument is something to be embraced. While it may be impossible get our arms around the whole elephant, the goal is to remain aware of its existence while at the same time truly describing the ear or leg before us.

References

Agar, M. (2006). An ethnography by any other name. *Forum Qualitative Social Research*, *7*(4), Art. 36. Available from http://www.qualitative-research.net/index.php/fqs/article/view/177

Baer, H. A., Singer, S., & Susser, I. (2003). *Medical anthropology and the world system*. Westport, CN: Praeger.

Becker, H. S. (1964). *Other side—Perspectives on deviance*. New York, NY: Free Press of Glencoe.

Becker, H. S. (1953). Becoming a marihuana user. *Americna Journal of Sociology, 59*(3), 235–242.

Berger, P.L., & Luckmann, T. ([1966] 1987). *The social construction of reality: A treatise in the sociology of knowledge*. Harmondsworth, UK: Penguin.

Bevan, M. T. (2014). A method of phenomenological interviewing. *Qualitative Health Research, 24*(1), 136–144.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research, 10,* 141–163.

Bourgois, P. (2002). Anthropology and epidemiology on drugs: The challenges of cross-methodological and theoretical dialogue. *International Journal of Drug Policy, 13,* 259–269.

Bourgois, P., & Schonberg, N. (2009). *Righteous Dope Fiend*. Berkeley, CA: University of California Press.

Buraway, M. (2014). Sociology as a combat sport. *Current Sociology, 62*(2), 140–155.

Carlson, R. G. (1999). 'Boy' and 'girl': The AIDS risk implications of heroin and cocaine symbolism among injection drug users. *Anthropology & Medicine, 6*(1), 59–77.

Carlson, R. G., Siegal, H. A., & Falck, R. S. (1995). Qualitative research methods in drug abuse and AIDS prevention research: An overview. In E. Y. Lambert, R. S. Ashery, & R. H. Needle (Eds.), *Qualitative methods in drug abuse and HIV research, NIDA research monograph 157*. National Institute on Drug Abuse: Rockville, MD.

Clark, L. (1971). *Tulsa*. New York, NY: Lustrum Press.

Clatts, M. C., Welle, D. L., Goldsamt, L. A., & Lankenau, S. E. (2002). An ethno-epidemiological model for the study of trends in illicit drug use: Reflections on the 'emergence' of crack injection. *International Journal of Drug Policy, 13*(4), 285–295.

Clatts, M. C., Sotheran, J., Heimer, R., & Goldsamt, L. (1999). Interdisciplinary research on transmission of bloodborne pathogens among drug injectors: Applications of ethnography in epidemiology and public health. In P. Marshall, M. Singer, & M. Clatts (Eds.), *Integrating cultural, observational, and epidemiological approaches in the prevention of drug abuse and HIV/AIDS*. NIDA: Rockville, MD.

Clatts, M. C., Giang, L. M., & Goldsamt, L. A. (2007). Novel heroin injection practices: Implications for transmission of HIV and other bloodborne pathogens. *American Journal of Preventive Medicine, 32*(6S), S226–S233.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography*. Berkely, CA: University of California Press.

Dai, B. (1937). *Opium addiction in Chicago*. Montclair, NJ: Patterson Smith.

Daniulatyte, R., Falck, R., Li, L., Nahhas, R. W., & Carlson, G. R. (2012). Respondent-driven sampling to recruit young adult non-medical users of pharmaceutical opioids: Problems and solutions. *Drug and Alcohol Dependence, 121*(1–2), 23–29.

Davenport-Hines, R. (2001). *The pursuit of oblivion: A global history of narcotics*. London, UK: W.W. Norton.

De Quincey, T. (2003). *Confessions of an english opium-eater and other writings*. London, UK: Penguin Classics.

Dreher, J. (2011). Alfred Schutz. In G. Ritzer & J. Stepnisky (Eds.), *The wiley-blackwell companion to major social theorists (volume 1)*. Malden, MA: Wiley-Blackwell.

Draus, P., Roddy, J., & Asabigi, K. (2015a). Making sense of the transition from the Detroit streets to drug treatment. *Qualitative Health Research, 25*(2), 228–240.

Draus, P., Roddy, J., & Asabigi, K. (2015b). Streets, strolls and spots: Sex work, drug use and social space in Detroit. *International Journal of Drug Policy, 26,* 453–460.

Draus, P., & Roddy, J. (2013). *Building a healthy community in detroit: Tracking the impact of the hope village initiative area*. In Institute Integrated Assessment Report Series Volume III Report 2. Ann Arbor, MI: University of Michigan, Graham Institute for Environmental Sustainability. Available at http://graham.umich.edu/media/files/HOPEVillageIA_Baseline.pdf.

Draus, P., Roddy, J., & Greenwald, M. (2012). Heroin mismatch in the motor city: Addiction, segregation and the geography of opportunity. *Journal of Ethnicity in Substance Abuse, 11*(2), 149–173.

Draus, P. J., Roddy, J., & Greenwald, M. (2010a). A hell of a life: Addiction and marginality in post-industrial detroit. *Social and Cultural Geography, 11*(7), 663–680.

Draus, P., Roddy, J., & Greenwald, M. (2010b). I always kept a job: Income generation, heroin use and economic uncertainty in 21st century detroit. *The Journal of Drug Issues, 40*(3), 841–870.

Draus, P., & Carlson, R. G. (2009). Down on main street: Drugs and the small town vortex. *Health & Place, 15* (1), 247–254.

Draus, P., & Carlson, R. G. (2007). Change in the scenery: An ethnographic exploration of crack cocaine use in rural Ohio. *Journal of Ethnicity in Substance Abuse, 6*(1), 81–107.

Draus, P., & Carlson, R. G. (2006). Needles in the haystacks: The social context of initiation to heroin injection in rural Ohio. *Substance Use and Misuse, 41* (8), 1111–1124.

Draus, P., Siegal, H. A., Carlson, R. G., Falck, R. S., & Wang, J. (2005). Cracking the cornfields: Recruiting illicit stimulant drug users in rural Ohio. *The Sociological Quarterly, 46,* 165–189.

Draus, P. (2004). *Consumed in the city: Observing tuberculosis at century's end*. Philadelphia, PA: Temple University Press.

Draus, P. (2000). Spying on an eyesore: Space, place and urban decay. In R. Hutchison (Ed.), *Constructions of Urban Space*. Stamford, CT: JAI Press.

Duff, C. (2011). Reassembling (social) contexts: New directions for a sociology of drugs. *International Journal of Drug Policy, 22,* 404–406.

Duneier, M. (1992). *Slim's table: Race, respectability, and masculinity*. Chicago, IL: The University of Chicago Press.

Galliher, J. F., Keys, D. P., & Elsner, M. (1998). Lindesmith v. Anslinger: An early government victory in the failed war on drugs. *Journal of Criminal Law and Criminology, 88*(2), 661–682.

Garfinkel, H. ([1967] 2003) *Studies in ethnomethodology*. Cambridge, MA: Polity Press.

Gobo, G. (2005). The renaissance of qualitative methods. *Forum: Qualitative Social Research, 6*(3), Art 42.

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Englewood Cliffs, NY: Prentice-Hall.

Golub, A., Johnson, B. D., & Dunlap, E. (2005). Subcultural evolution and illicit drug use. *Addiction Research & Theory, 13*(3), 217–229.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*(2), 174–199.

Henderson, D. J., Boyd, C., & Mieczkowski, T. (1994). Gender, relationships and crack cocaine: A content analysis. *Nursing and Health, 17,* 265–272.

Hillman, D. C. A. (2008). *The chemical muse: Drug use and the roots of western civilization*. New York, NY: Thomas Dunne Books.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *The Sociology of Health and Illness, 16*(1), 103–121.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. New York, NY: Oxford University Press.

Law, J. (2009). Actor network theory and material semiotics. In B. S. Turner (Ed.), *The new blackwell companion to social theory*. Wiley-Blackwell: West Sussex, UK.

Lindesmith, A. R. (1968). *Addiction and Opiates*. Chicago, IL: Aldine.

Lindesmith, A. R. (1965). *The addict and the law*. Bloomington, IN: Indiana University Press.

McKnight, C., Des Jarlais, D., Bramson, H., Tower, L., Abdul-Quader, A. S., Nemeth, C., et al. (2006). Respondent-driven sampling in a study of drug users in New York City: Notes from the field. *Journal of Urban Health, 83*(7), i54–i59.

Moore, D. (1993). Ethnography and illicit drug use: Dispatches from an anthropologist in the "field". *Addiction Research, 1,* 11–25.

Neaigus, A., Friedman, S. R., Curtis, R., Des Jarlais, D. C., Furst, R. T., Jose, B., et al. (1994). The relevance of drug injectors' social networks and risk networks for understanding and preventing HIV infection. *Social Science and Medicine, 38,* 67–78.

Needle, R. H., Genser, S. G., & Trotter, R. T. (Eds.). (1995). *Social networks, drug abuse, and HIV transmission. NISA research monograph 151*. Rockville, MD: National Institute on Drug Abuse.

Nicholson, T., White, J., & Duncan, D. (1998). Drugnet: A pilot study of adult recreational drug use via the WWW. *Substance Abuse, 19*(3), 109–121.

Nichter, M., Quintero, G., Nichter, M., Mock, J., & Shakib, S. (2004). Qualitative research: Contributions to the study of drug use, drug abuse, and drug use(r)-related interventions. *Substance Use and Misuse, 39,* 1907–1969.

Palmer, B. D. (2000). *Culture of darkness: Night travels in the histories of transgression*. New York, NY: Monthly Review Press.

Page, J. B., & Singer, M. (2010). *Comprehending drug use: Ethnographic research at the social margins*. New Brunswick, NJ: Rutgers University Press.

Parker, H., Williams, L., & Aldridge, J. (2002). The normalisation of 'sensible' recreational drug use: Further evidence from the North West Longitudinal Study. *Sociology, 36*(4), 941–964.

Parker, H., Measham, F., & Aldridge, J. (1998). *Illegal leisure: The normalisation of adolescent recreational drug use*. London, UK: Routledge.

Primack, B. A., Dalton, M. A., Carroll, M. V., Agarwai, A. A., & Fine, M. J. (2008). Content analysis of tobacco, alcohol, and other drugs in popular music. *Archives of Pediatric and Adolescent Medicine, 162* (2), 169–175.

Quintero, G. A., & Estrada, A. L. (1998). Cultural models of masculinity and drug use: "Machismo", heroin, and street survival on the U.S.-Mexico border. *Contemporary Drug Problems, 25*(1), 147.

Roddy, J. & Draus, P. (2014, November). *Given the greenlight in Detroit: Photographic illustrations of themes during medical marijuana research*. In Paper presented at the American Public Health Association 142nd Annual Meeting, New Orleans, LA.

Schutz, A. (1967). *The phenomenology of the social world*. Evanston, IL: Northwestern University Press.

Scott, J. (2013). *Social network analysis* (3rd ed.). Los Angeles, CA: Sage.

Sevigny, E. L., Pacula, R. L., & Heaton, P. (2014). The effects of medical marijuana laws on potency. *International Journal of Drug Policy, 25,* 308–319.

Shaw, C. R. (1930). *The Jack-Roller: A delinquent boy's own story*. Chicago, IL: The University of Chicago Press.

Singer, M. (2013). Studying hidden and hard-to-reach populations. In J. J. Schensul & M. D. LeCompte (Eds.), *Specialized ethnographic methods: A mixed methods approach*. Lanham, MD: AltaMira Press.

Smith, M. P. (1992). Postmodernism, urban ethnography, and the new social space of ethnic identity. *Theory and Society, 21,* 493–531.

Snell, P. (2010). From Durkheim to the Chicago school: Against the 'variables sociology' paradigm. *Journal of Classical Sociology, 10,* 51–67.

Spradley, J. P. 1970 (2000). *You Owe Yourself a Drunk: An Ethnography of Urban Nomads*. Prospect Heights, IL: Waveland Press.

Spradley, J. P. (1980). *Participant observation*. New York, NY: Holt, Rinehart and Winston.

Spradley, J. P. (1979). *The ethnographic interview*. New York, NY: Harcourt, Brace, Jovanovich.

Thomas, W. I., & Thomas, D. S. (1928). *The child in America: Behavior problems and programs*. New York, NY: Knopf.

Thomas, W. I., & Znaniecki, F. (1958 [1918/1920]). *The polish peasant in Europe and America, Volumes I and II*. New York, NY: Dover Publications.

Trotter, R. T., Bowen, A. M., & Potter, Jr., J. M. (1995). Network models for HIV outreach and prevention programs for drug users. In R. H. Needle, S. G. Genser & R. T. Trotter (Eds.), *Social networks, drug abuse, and HIV transmission*. NISA Research Monotraph 151. Rockville, MD: National Institute on Drug Abuse.

Valdez, A., & Kaplan, C. D. (1998). Reducing selection bias in the use of focus groups to investigate hidden populations: The case of Mexican-American gang members from South Texas. *Drugs & Society, 14*(1–2), 209–224.

Valente, T. W., Gallaher, P., & Mouttapa, M. (2004). Using social networks to understand and prevent substance use: A transdisciplinary perspective. *Substance Use and Misuse, 39*(10–12), 1685–1712.

Valdez, A., Neaigus, A., & Kaplan, C. D. (2008). The influence of family and peer risk networks on drug use practices and other risks among Mexican American non-injecting heroin users. *Journal of Contemporary Ethnography, 37*(1), 79–107.

Wang, C. C. (1999). Photovoice: A participatory action research strategy applied to women's health. *Journal of Women's Health, 8*(2), 185–192.

Wang, C. C., & Burris, M. A. (1997). Photovoice: Concept, methodology, and use for participatory needs assessment. *Health Education & Behavior, 24*(3), 369–387.

Watters, J. K., & Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. *Social Problems, 36*(4), 416–430.

Weeks, M. R., Clair, S., Borgatti, S. P., Radda, K., & Schensul, J. J. (2002). Social networks of drug users in high-risk sites: Finding the connections. *AIDS and Behavior, 6*(2), 193–206.

Whyte, W. F. (1943). *The street corner society: The social construction of an Italian Slum*. Chicago, IL: The University of Chicago Press.

Zinberg, N. E. (1984). *Drug, set, and setting: The basis for controlled intoxicant use*. New Haven, CT: Yale University Press.

Qualitative Data Analysis in Drug Research

8

Henry H. Brownstein

8.1 Introduction

The methods of social scientific research enable us to conceptualize and analyze our experience of social life, in order to describe, understand, and explain it as it is presented to us, in a form we recognize as empirical reality (Kaplan 1964; Lazarsfeld and Rosenberg 1955). To accomplish this end, different methods are used in different ways depending on things such as the subject of interest, what questions are being asked, and how the researcher defines social reality.

Despite contemporary interest in mixed methods (Bazeley 2009; Creswell 2003; Small 2011), and despite the fact that both approaches "claim to reveal some truth about the world" (Martin and Stenner 2004), the sharpest dichotomy exists between quantitative and qualitative methods. Conceptually, they differ in that quantitative studies attend to social phenomena as objects and emphasize experimentation in order to eliminate plausible explanations, while qualitative studies attend to social phenomena as subjects and emphasize methods that seek meaning in symbolic representations, such as words and images. Methodologically, they differ in that quantitative methods are designed for the study of relationships among discrete, precisely defined and measured variables, while qualitative methods are designed for the study of commonalities among broadly conceptualized social phenomena (Ragin and Amoroso 2011).

By design, qualitative methods are used to "produce descriptive data [from] people's own written or spoken words and observable behavior" (Bogdan and Taylor 1975). In that sense, data collected using a qualitative approach include methods such as ethnographic observation and open-ended interviewing. These methods allow the researcher to gain a deeper understanding of the meaning of an instance of social reality as a subjective experience from the perspective of its participants (Bogdan and Taylor 1975; Brownstein 1983, 1990; Denzin and Lincoln 1994; Gubrium and Holstein 1997; Hesse-Biber and Leavy 2004; Warren and Karner 2010). This chapter is about the application and significance of qualitative methods for the study of people living in society who are involved with illicit drug use and trade.

8.2 The Study of Illicit Drug Use and Trade

In contemporary societies, drug involvement in one way or another has been established under a socially recognized, legitimate authority to be unlawful. When something is unlawful, such as the use of or trade in drugs, public records of transactions, compliance with government

H.H. Brownstein (✉)
Center for Public Policy, The Wilder School for Government and Public Affairs, Virginia Commonwealth University, 921 West Franklin Street, Richmond, VA, USA
e-mail: hhbrownstein@vcu.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_8

regulations, and official documentation of business activity, including items such as earnings, revenue, costs, or tax liability, are unavailable. Moreover, there are neither official records of the number of individuals and/or organizations engaged in the business of buying or selling these drugs, nor a number of consumers or patterns of consumption. Thus, an official accounting of the number of people engaged in the use, or trading of, illicit drugs is unavailable for business and research purposes (Brownstein 2000).

Without adequate record of the illicit drug industry, local drug markets and the people involved in them are largely hidden from public scrutiny and study (Lambert 1990). Without official documentation of who is involved and how things are organized, it is, at best, a challenging notion to construct a trustworthy statistical accounting, or to conduct a compelling statistical analysis of the scope, organization, operation, characteristics, practices, producers, distributors, and/or consumers of the illicit drug industry. Policymakers who need to make decisions to address problems associated with society's illicit drug involvement, in addition to the researchers who use quantitative approaches to study illicit drug involvement, have used surrogate measures as proxies to approximate what could not otherwise be measured or counted (Thoumi 2005).

To measure the scope and magnitude of illicit drug consumption, economists, social scientists, and epidemiologists have collected and analyzed surrogate measures based on regional and national population surveys. For example, U.S. data from the National Survey of Drug Use and Health (NSDUH), a nationwide household survey conducted annually for the federal government by the Substance Abuse and Mental Health Services Administration (SAMHSA), has been utilized. A random sample of 70,000 respondents, ages 12 and older, are surveyed about their use of and experience with various illicit drugs (SAMHSA 2014). Other SAMHSA surveys collecting public health data related to drug use have also been useful for quantitative analyses and reporting, such as the Treatment Episode Data Set (TEDS), which collects data about drug treatment and use for people admitted to drug treatment in a given year (SAMHSA 2012), and the Drug Abuse Warning Network (DAWN) program, which actively collects demographic and related health data from Emergency Department (ED) admissions that were the consequence of either licit or illicit substance misuse or abuse (SAMHSA 2013). Focusing on a subpopulation known for a disproportionate number of heavy drug users, the National Institute of Justice (NIJ), of the US Department of Justice, and the Office of National Drug Control Policy (ONDCP), of the Office of the President, separately sponsor a survey to collect interview data and urine samples from arrestees shortly after arrest, in order to measure the extent to which they had been using one of the ten different illicit substances relative to the time of arrest (NIJ 2003; ONDCP 2014).

To measure the scope and magnitude of production and distribution in the illicit drug industry, policymakers, epidemiologists, and other researchers have used data from public agencies that respond to events that are related to the involvement of people with illicit drugs, particularly the public safety responses of law enforcement. A major source of the number of events and people involved with illicit drugs is the Uniform Crime Reports (UCR) produced by the Federal Bureau of Investigation (FBI), a yearly data collection from law enforcement agencies across the country, voluntarily and without verification, that reports crime rate, crime type (not including drug crimes), and overall arrest count (FBI 2014). A criminal justice data collection more directly designed to estimate the quantity of illicit drugs available for distribution is the System to Retrieve Information from Drug Evidence (STRIDE) program, sponsored by the Drug Enforcement Administration (DEA), which collects data and produces reports based on laboratory analyses of drugs seized by law enforcement agencies and later submitted to the DEA for analysis (NDIC 2014).

As noted earlier, quantitative analyses are designed to study relationships among discrete and precisely defined and measured variables. When measuring for a surrogate of another measure, thereafter performing a quantitative

analysis that informs the question (or questions) you are asking is challenging at best; however, and especially in the case of hidden populations, surrogate measures are sometimes all one has to work with. Qualitative analyses, as noted above, seeks to produce meaning from symbolic representations, such as words and images, and are designed to study commonalities across broadly conceptualized social phenomena.

The foundation of qualitative research is the discovery of grounded theory, which in lay terms refers to the generation of theory from data. In reference to grounded theory, Glaser and Strauss (2012/1967) argue for qualitative research methods that emphasize the discovery of theory grounded in observations of natural settings. In this sense, qualitative methods have been particularly useful and productive for the generation of theories about the lives and experience of people, such as hidden populations involved with illicit drugs. Using qualitative methods of data collection and analysis, researchers have been able to learn about things they might not otherwise be able to see or measure. By themselves, for example, they can contribute to our understanding of the organization and operation of illicit drug markets, in addition to those who participate in them. They can also contribute to our capacity for theory construction, which can then be used to generate hypotheses, conceptualize observed variance, and discuss how such items might best be measured in quantitative or epidemiological analysis.

In order for qualitative studies to be of significant value, subsequent findings must be trustworthy, in that they are credible, dependable, and confirmable (Lincoln and Guba 1985). Trustworthiness, with respect to qualitative research, is a function of subjective adequacy, and relies heavily upon whether the researcher is able to establish, to the satisfaction of others, a conceptual construction of the social phenomenon that is recognized and understood by the social actors who participate in that phenomenon in real-life experience (Altheide and Johnson 1994; Brownstein 1983; Schutz 1954; Weber 1947). For researchers studying those involved with illicit drugs, careful attention to issues of sampling, data collection, and data analysis are necessary.

8.3 Qualitative Methods in the Study of Illicit Drug Involvement

The early twenty-first century is a period of renewed interest in the application of qualitative research methods to the study of drugs in society (Rhodes and Moore 2001), but the history of this approach, with respect to the study of illicit drug involvement, goes back almost 200 years. Arguably the earliest qualitative examination of drug use was written in the early nineteenth century by Thomas De Quincey, well before drugs were relegated to a world of illegitimacy, as an autobiographical account of his experiences with opium (De Quincey and Morrison 2013). By the early twentieth century, the use of qualitative methods to study societal drug use and involvement had gained prominence, most notably in Chicago and other major U.S. cities, as well as the UK, Australia, and elsewhere (Rhodes and Moore 2001).

The following sections discuss the types of sampling, data collection, and data analysis used by qualitative researchers studying illicit drug involvement, and includes particular studies in this area, in order to describe how qualitative studies have been designed, as well as how they have contributed to our knowledge and understanding of drugs in society.

8.3.1 Sampling

People involved as producers, distributors, and/or consumers of illicit drugs, for the most part, are hidden and out of reach to public policymakers, drug treatment providers, and researchers alike. In a report written for the National Institute on Drug Abuse (NIDA), Lambert and Wiebel (1990) refer to this hidden population as those involved with illicit drugs who are: disadvantaged and disenfranchised, homeless, transient, mentally ill, high school

dropouts, criminal offenders, prostitutes, juvenile delinquents, gang members, runaways, and other 'street people'; little information is available on these populations, despite the general awareness of their presence. As such, these populations are often omitted from nationally representative surveys, largely because they lack a fixed address, are less likely to be found at home, or otherwise refuse to interview. Considering the aforementioned characteristics, it should be of no surprise that those who belong to hidden populations are at greater risk for drug abuse and drug-related morbidity than the general population. Ironically, the very individuals who would benefit from drug abuse treatment and prevention efforts are the least studied, the least understood, and the most elusive to epidemiologists, clinicians, researchers, and others concerned with understanding and improving the public health of these populations (Lambert and Wiebel 1990).

With this in mind, Wiebel (1990) suggests that "…the nature of substance abuse in our society, together with applied research priorities, appears to be of greatest significance in maintaining a rich field of inquiry for exploratory and descriptive scientific investigations" for which he identifies two main reasons. First, the patterns, trends, and substances of interest amongst users, abusers, and distributors change so rapidly that there are constantly new drugs available and new problems to deal with, lending considerable difficulty for consistent observation and measurement. Second, policymakers and practitioners need reliable, up-to-date information from which to make informed and pragmatic decisions, and "…qualitative research is often the only appropriate means available for gathering sensitive and valid data from otherwise elusive populations of substance abusers" (Wiebel 1990).

Upon recognizing the value of qualitative research methods to study illicit drugs in society, the first challenge is to establish scientifically sound ways to identify appropriate hidden populations and thereafter collect data from them. Early studies used a method known as 'snowball', or 'chain referral sampling', an approach that "yields a study sample through referrals made among people who share or know others who possess some characteristics that are of research interest" (Biernacki and Waldorf 1981). Snowball sampling has been widely used, given that it is convenient, economical, efficient, and effective (Sadler et al. 2010; van Meter 1990). Snowball sampling can be problematic, however, with potential biases due to (1) the fact that it is a non-probability method of sampling, (samples tend to be unbalanced and favor selected demographic characteristics), and (2) because there is no way to know when a sufficient sample size or representativeness has been reached (Sadler et al. 2010).

In more recent years, efforts have been underway to design sampling strategies that build on the snowball model but mathematically address some of these aforementioned concerns. Most notably, researchers have developed an approach known as respondent-driven sampling (RDS) to recruit samples from hidden populations (Gile and Handcock 2010; Heckathorn 1997, 2007). RDS is a new form of chain-referral sampling and, as explained by Heckathorn (1997), is based "on an analysis drawing on Markov chains and the theory of biased networks [showing] that suitable incentives can reduce the biases of chain-referral samples." He goes on to stress that there are limitations to RDS, as it is only suitable in particular circumstances, such as "for sampling populations with a contact pattern [wherein] the activities that constitute membership in the population must create connections among population members, as when drug users purchase or share drugs" (Heckathorn 1997). Consequently, while it has not been easy to meet these standards, qualitative drug researchers concerned with sampling bias have shown interest in the application of RDS to their research.

8.3.2 Data Collection

Ethnography and open-ended interviewing are two common methods of collecting data in qualitative analysis to study those populations who supply and/or consume illicit drugs, in addition to the experience of those people in

society. While they do overlap, and it is not uncommon for a researcher to utilize elements of each method, ethnography and open-ended interviewing are distinctive approaches to data collection. Ethnography refers to the participation of a researcher "overtly or covertly, in people's daily lives for an extended period of time, watching what happens, listening to what is said, asking questions; in fact, collecting whatever data are available to shed light on the issues with which he or she is concerned" (Hammersley and Atkinson 1983). Open-ended interviews are a distinctive method, designed to gain in-depth information and understanding about the knowledge, awareness, and perspectives of participants in a given social setting, incident, or experience. They can be conducted as an informal conversational interview with limited constraints (a useful approach for ethnographers asking questions), a more structured interview allowing the respondent to say what they want but directed by the interviewer following an outline or set of topics, or an even more structured interview with specific questions that allows the respondent to say what he or she wants to say without constraint.[1]

Ethnography has been widely used to study drugs in communities (Murphy et al. 2016) and the continuing focus on particular communities has been a good fit for these studies. As participant observers, those who perform this kind of research, at some level, have to become engaged with the community they are studying; once the necessary rapport and relationships are established between researchers and key community informants, researchers have an incentive to maintain those relationships and continue the work. This is particularly valuable when the research focus is a hidden population, and the rapport and relationships established open an ongoing opportunity to study that population. Throughout the twentieth century and to the present day, a number of communities across the world have become centers for ethnographic research, focusing on the involvement of people with illicit drugs. There have been studies of those who use illicit drugs, patterns of illicit drug use, local drug transactions, and drug markets. In the event of a public safety concern or health crisis, such as violence related to the use of illicit drugs (or licit drugs in illicit ways), ethnographic studies have been useful to study and understand the related problems for people and their communities. The following examples are a mere fraction of the ethnographies conducted in the US to study drugs.

Arguably, the ethnographic tradition of research to study hidden populations took hold in the U.S. early in the twentieth century through the work of sociologists in Chicago (Jaynes et al. 2009). Researchers at the University of Chicago conducted a number of studies of life in local communities to understand the relationship between the experience of the people living in the community and the ecological characteristics of their surroundings, many studying hidden populations of people who lived outside of the law, often looking at the experiences through the eyes of a key informant (Shaw 1930; Thrasher 1927; Zorbaugh 1929). Demonstrating the value of this research for understanding problems associating with drug use, Alfred Lindesmith, a researcher from the University of Chicago, conducted ethnographic research in the Chicago area by way of in-depth interviews with opium users. His findings addressed the contemporary lack of knowledge regarding the causes of, and appropriate treatment for drug addiction, when he concluded that "deprivation is the essential factor both in the origin of the craving and in its growth" (Lindesmith 1938). Decades later, Patrick Hughes and colleagues conducted an epidemiological ethnographic study of heroin addicts in Chicago, from which he observed what he referred to as an 'addict subculture', in which participants, users, and dealers were engaged in some form of violence on a daily basis (1977). Touching on these findings, Hughes noted that "Deviants of all sorts frequented the addicts'

[1]Compare to the practical guide for qualitative interview design for novice investigators prepared by Turner (2010). This guide explores interview protocols and sample questions appropriate across three distinctive categories of qualitative study design: (1) informal conversational interviews; (2) general interview guide approaches; and (3) standardized open-ended interviews.

street hangouts, and in these high-crime neighborhoods there was always the risk of confrontation by a drunk, or by members of a delinquent gang who wanted to take away an addicts' freshly stolen television set" (1977).

A strong ethnographic research tradition surrounding the study of drug use and drug markets also developed in New York City by the mid-twentieth century, and New York became a center for ethnographic drug research in the years that followed. This is not surprising given the growing political and media concern about drugs during that period, particularly heroin, cocaine, and later crack cocaine (Brownstein 2013; Inciardi 1992). Contrary to popular belief at the time that heroin abusers were incapable of functioning in society, Edward Preble, an anthropologist, and his colleague John Casey conducted ethnographic observations and open-ended interviews with addicts on the streets of New York, from which they concluded that "Their behavior is anything but an escape from life. They are actively engaged in meaningful activities and relationships seven days a week" (Preble and Casey 1969). They described the daily life of a heroin addict with a phrase used by their subjects; the addicts, they said, were "taking care of business" (Preble and Casey 1969). Preble's influence on drug research in New York demonstrates how the rapport and relationships built in an earlier study can lead to ongoing ethnographic work in an area, and potentially even careers for younger colleagues. In 1975, Preble moved to Narcotic and Drug Research, Inc. (NDRI), a research organization in the city that was becoming known for its study of both people and communities involved with illicit drugs. While at NDRI, Preble pioneered and introduced to his colleagues an ethnographic method of collecting data on hidden populations, which is often referred to as the 'storefront methodology' (Johnson et al. 1985). Using the storefront method, a research team locates, opens, and staffs a storefront in a neighborhood where heroin is a known endemic, and there employs ex-addicts and ex-offenders as field workers to recruit active heroin users in the neighborhood to be interviewed at the storefront (Johnson et al. 1985). Bruce Johnson, who worked with and learned from Preble at NDRI, utilized this method to study the economic behavior of heroin users in the East and Central Harlem neighborhoods of the city (Johnson et al. 1985). Later, Paul Goldstein, also a researcher who worked with and learned from Preble, used the same method to conduct a study on the Lower East Side of New York to examine the relationship between drug use and violence (Goldstein et al. 1990). Given the fertile ground for ethnographic drug research in New York, other researchers have also independently conducted such research in the city.[2]

During the 1960s, drugs were a large part of a countercultural movement that took place in cities around the world and on a number of levels. This movement was comprised largely of young people who believed that anything was possible, and was centered in San Francisco (Gitlin 1993; Roszak 1968). As such, San Francisco proved to be a good place for ethnographic studies of drug use and drug transactions. Building on the ethnographic tradition of drug communities studies in Chicago, Dan Waldorf studied the lives of drug users in San Francisco in the context of a career as opposed to deviant behavior, describing drug use as an orderly and purposeful way to survive daily experience, with discernable stages and transitions (1973). Over the years, he conducted ethnographies of drug users and drug use with colleagues at the Institute for Scientific Analysis (ISA), wherein he continued to study drug users as people living out their lives (Waldorf et al. 1992). As a product of

[2]Examples include Terry Williams' (1992) classic ethnographic study of the economics and community of residents in a crackhouse in Manhattan's West Spanish Harlem, Maher's (2000) detailed 3-year ethnographic study of the economic lives of women drug users in New York City, focusing on divisions of labor in the street-level economy, and Bourgois' (2003) participant observational study of poverty and social marginalization in inner-city America as experienced by street-level drug dealers in East Harlem. All three works draw extensively on their subjects own words to graphically depict the lived experiences of those trapped in urban poverty; enriching the understanding of the social/psychological roots of the drug problems they suffer.

this approach, a distinct line of ethnographic research has developed, which continues to look at the lives of those whose experience and efforts to live their lives as drug users is complicated by the fact that they are in some way social outsiders.[3] These lines of ethnographic research continue in San Francisco to the present day.[4]

As qualitative methods of collecting data, both ethnography and open-ended interviewing are well suited to research that emphasizes the discovery of grounded theory through observations of natural settings or experience (Glaser and Strauss 2012/1967). In this regard, ethnography has been a very successful method for studying a community through a particular group of people, over an extended period of time. Alternatively, open-ended interviewing is preferred when the focus of study is a particular social phenomenon, as identified from the perspectives of an expanding number of people.

In the tradition of grounded theory, Paul Goldstein, a New York City drug researcher utilizing the storefront ethnographic method to study the relationship between drugs and violence, conceptualized that drugs and violence could be related in three different ways (1985). Goldstein proposed what he called a 'tripartite framework', suggesting that drugs and violence could (1) be the product of a psychopharmacological reaction produced within a person using a particular drug, (2) the outcome of an unsuccessful interpersonal drug transaction (such as drug dealers fighting over territory or a drug buyer feeling cheated), or (3) the consequence of the compulsive need of a drug user to get drugs or money to satisfy that need. The tripartite framework produced a great deal of interest among researchers, and Goldstein and his colleagues in the years that followed conducted a number of studies collecting empirical data to test the hypothesis. The challenge of such a framework was not to focus on a single community of drug users, but rather to reach as many people as possible who had both used drugs and committed a violent act. They began their search for empirical evidence of the relationship between drugs and violence with a study in New York State, using semi-structured open-ended interviews to ask those in state custody (in state facilities) for having committed a homicide about the involvement of drugs in the homicide and, more broadly, in their lives (Brownstein et al. 1992; Goldstein et al. 1989, 1992; Spunt et al. 1994). This study illustrated the great extent to which drugs and violence had been involved in the lives of the people who ultimately killed another person, but it also affirmed that when drugs were directly related to a homicide event (and they often were), different drugs were related in different ways. These findings ultimately produced more questions than answers and thereby resulted in a series of studies using open-ended interviews to study the relationship between violence and drugs. For example, one subsequent study involved interviews administered only to women in prison for murder and found that the drugs in their lives, and their involvement in the murder, was different from that of the men interviewed in the earlier study (Brownstein et al. 1995; Crimmins et al. 1997; Spunt et al. 1996). Studies in New York and Maryland conducted with youth in custody collected data from open-ended interviews about drugs and violence in their experience and similarly found inextricable, but not necessarily direct connections between drug

[3]See for example, Hunt et al.'s (1997) use of snowball sampling and in-depth interviews of Southeast Asian gang members to study culture and ethnic identity among Asian gang members in Northern California. This study used a three-stage process. In the first stage, respondents answered questions from a qualitative life history schedule, which explored the problems individuals had experienced in their country of origin and in the U.S. during the process of migration. The second stage utilized a quantitative interview schedule addressing topics such as basic demographics, drug use, and gang related activity. The third stage was comprised of in-depth interviews exploring topics such as gang history and criminal activity.

[4]Examples include Maloney et al.'s (2009) use of qualitative interview data to examine fatherhood as a potential turning point in the lives of gang members, Murphy and Rosenbaum's (1999) in-depth interviews with over 120 women who had children while using drugs, and Sales and Murphy's (2007) use of qualitative in-depth interviews to study the motivations and circumstances associated with ecstasy distribution in San Francisco.

involvement and violence (Crimmins et al. 2000; Ryder et al. 2009).

Campbell collected drug trafficking stories from what might be referred to as conversational interviews conducted "in family contexts, in parties, at work, in chance meetings on the street or businesses, and from neighbourhoods in which [he] resided" over ten years while in a community near the border between the United States and Mexico (2005:328). He used the interviews to collect stories told amongst people living near the border about the folklore and culture of drug trafficking in the area. He used the stories "not to deconstruct them in order to cast blame on the storytellers or to frame them as valiant utterances of resistance," but rather to "illustrate a cultural process" (Campbell 2005). From his analysis of these stories and how they are used, he concluded that "for the border population, drug trafficking is a tacitly tolerated activity or a mundane, everyday phenomenon that, though not fully accepted, is not considered a radically deviant or unusual lifestyle" (2005:333).

In a national study of methamphetamine markets, open-ended interviews were conducted with people who participated in the markets in different parts of the US. The research team visited more than 28 cities and towns in 5 regions of the country (Southeast, Middle Atlantic, Midwest, Southwest, and Pacific Northwest), where they observed and talked with local police, as well as local and regional public health and safety officials, drug treatment and prevention workers, family service providers, methamphetamine users, sellers, cooks, and other people who knew about the local meth markets in their community and region (Brownstein et al. 2014). Using semi-structured and conversational interviews, they collected data about the organization and operation of local markets, how they related to regional markets and the national industry, and the impact of changing state and federal policies and laws on the local markets and users (Brownstein et al. 2014).

In some studies, open-ended interviewing was used as part of a larger ethnographic study, focusing particularly on a key informant. For example, Sudhir Venkatesh studied urban poverty, gangs, and drugs by embedding himself in a housing project in Chicago and forming a close relationship with a gang leader who ran a business selling crack, allowing him to learn from ongoing conversations with his informant and to informally interview others in the community (Venkatesh 2008). In New York, Philippe Bourgois began with a plan to study an urban underground economy and ended up studying a community of street-level crack dealers in East Harlem (1996). This study was designed as an ethnography, and much of what he learned came from personal conversations and communications with a small group of people he got to know in the Spanish Harlem crackhouses (Bourgois 1996).

Open-ended interviews have also been used to study the impact of public policy on drug using populations within in the context of public health concerns or crises. Stephen Koester (1994) used open-ended interviews when he studied syringe sharing among injection drug users in Denver. As part of an ethnographic study, he used both "formal interviews and casual conversations" to collect data about "everyday activities in which street-based injectors engage, and included detailed discussions about syringe sharing, drug coping, personal economic strategies, and the impact of law enforcement" (Koester 1994). Koester found that the risky practices of injection drug users who shared needles and injected in "high risk environments," such as shooting galleries, was not the "maladaptive rituals of a vast drug subculture," but rather a response to the management of risk among people who were more fearful of being arrested for criminal possession of a syringe than being infected with HIV by sharing needles (1994). After Hurricane Katrina created a public health and safety crisis in New Orleans, Eloise Dunlap and Andrew Golub (2011) conducted semi-structured, open-ended interviews with surviving poor, as well as African-American drug users and sellers, to learn how and why these users and dealers, who were more concerned with continuing their drug use and business than they were with their own safety, continued to party during and loot after the storm rather than evacuate.

8.3.3 Data Analysis

Whether data is collected from an ethnographic study or from open-ended interviewing, what can be done to make sense of all the narrative and symbolic data collected by the qualitative methods of the data collection? Over the years, qualitative researchers, such as those studying drug use and the drug trade, have developed methods for data analysis. With that being said, it is important to note that qualitative analysis is not like quantitative analysis. As noted earlier, quantitative methods collect data that can be analyzed to establish relationships among discrete and precisely measured variables, whereas qualitative methods collect data that can be analyzed to contribute to understanding the meanings of conceptualized social phenomena (Ragin and Amoroso 2011). As such, quantitative analysis techniques emphasizing internal validity, and being designed to eliminate plausible explanations (Campbell and Stanley 1963), are not appropriate for qualitative analysis. Rather, qualitative analysis emphasizes the need for methods that reveal meaning (Bogdan and Taylor 1975; Denzin and Lincoln 1994; Glaser and Strauss 2012/1967) and demonstrate construct validity in the broad sense that there is a connection between theoretical terms or constructs and measurements based on direct phenomena observation (Cherryholmes 1988; Messick 1975).

Based on philosophical traditions going back generations, analytic induction has been a foundation of qualitative analysis since early in the twentieth century, when it was described in general by Florian Znaniecki (1934), applied by Edwin Sutherland to the study of crime (1939), and utilized for drug research by Alfred Lindesmith in his study of opiate addiction (1938). Unlike deductive reasoning, which begins with a general hypothetical statement and conducts analyses of empirical data to reach a logical conclusion (Creswell 2003), analytic induction refers to the method of focusing on cases representing a broad social phenomenon and systematically analyzing them for commonalities toward a deeper conceptual understanding of the phenomenon (Ragin and Amoroso 2011; Robinson 1951; Turner 1953).

In his study of opiate addiction, Lindesmith used analytic induction and interpretation through an approach that looked for decisive negative evidence when comparing evidence from a number of cases representing a particular phenomenon: drug addiction (1938, 1947). He conducted personal interviews with more than 60 known addicts, and examined the available literature on other scholarly studies to test a theory of addiction based on the understanding or stressors related to addiction. In terms of his methodology, Lindesmith emphasized that it is "significant that the theory advanced in this study is not quantitative in form, nor is it a purely intuitive generalization which is not subject to proof, but that it is experimental in form in spite of the fact that it is based upon the analysis of data secured largely in personal interviews" (1938). This approach was based on considerations of the theory and testing its subjective adequacy with what he learned from his interviews, allowing for "the possibility of its own continuous reconstruction and refinement in terms of more extended experience and of more elaborated instances" (1938). Furthermore, and as described by Lindesmith, this approach searched for (and used) negative cases to examine and compare "succeeding tentative formulation (s)," each building on preceding formulations in the face of new evidence so that the "eventual hypothesis altered the preceding formulations sufficiently to include the cases which earlier had appeared as exceptions to the theory postulated" (1947).

Other drug researchers, using qualitative methods of data collection and analysis, have at least some extent of generally applied methods of analytic induction and interpretation. For example, Joseph Gusfield used a similar approach for his study of drinking and driving in California, when he began by looking at the established theory that the problem was a moral one, with the drivers being the source and used data from his study of police, courts, and others to find negative evidence that resulted in a new theoretical formulation arguing that the source of the

problem was the culture and the social institutions around drinking (1986). Patricia Adler, in a study of drug dealers, used open-ended interviews and learned about their entrepreneurial spirit and the rational order of the work they did against the backdrop of a theoretical perspective of drug dealers, as living and working in a subculture of hedonism (1993). A study involving open-ended interviews with women in prison for homicide tested the theory of the period that women who committed homicide typically did so in self-defense, killing men with whom they shared an abusive relationship (Rasche 1990), by examining cases in which women killed in the business context of a drug market transaction, to look for evidence that it was the transaction, and not an interpersonal relationship, that explained the homicide (Brownstein et al. 1995). In a more recent study involving mephedrone users in Northern Ireland, semi-structured interviews were conducted with 23 adults who had used the drug, and analytic induction was used (involving corroboration and comparison) to identify and explain outliers, resulting in a finding that initiation to use was influenced by market factors (McElrath and O'Neill 2011). Notably, while there is a large body of qualitative research on drugs that has conceivably been used the established methods of data collection, beyond Lindesmith's early example (and including the more recent examples cited here), there are not many that systematically have applied the method of analytic induction.

8.4 Conclusion

The history of drug research makes clear the importance of qualitative research as a contributor to our contemporary understanding of drug use, the drug trade, and the relationship between drugs and society (Lambert and Wiebel 1990; Martin and Stenner 2004; Rhodes and Moore 2001). While the methods of sampling, data collection, and data analysis for qualitative research have all been elaborated and developed to varying degrees, arguably more attention among drug researchers has been given to the sampling and collection of data, and less to the actual analysis (Adler 1990). Recognizing thoughtful and systematic analytic induction as a necessary method for trustworthy and significant qualitative research, Peter Adler wrote, "Induction in its present form allows us to call upon our experiences, utilize our feelings, and be a part of the research instrument while evolving and formulating theory that is grounded in the experiences of ourselves and the people we study. By remaining open and flexible, with a close eye and ear to behavior in natural settings, induction (or retroduction), is the only epistemology that addresses the subjective nature of human life, while generating theories that respect the everyday realities of its members" (1990).

References

Adler, P. (1990). Requirements for inductive analysis. In E. Y. Lambert (Ed.), *The collection and interpretation of data from hidden populations*. NIDA Research Monographs. Document Publication Number (ADM) 90-1678. Rockville, MD: National Institute on Drug Abuse.

Adler, P. A. (1993). *Wheeling and dealing: An ethnography of an upper-level drug dealing and smuggling community* (2nd ed.). New York, NY: Columbia University Press.

Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Bazeley, P. (2009). Editorial: Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research, 3,* 203–207.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research, 10*(2), 141–163.

Bogdan, R., & Taylor, S. (1975). *Introduction to qualitative research methods—A phenomenological approach to the social sciences*. New York, NY: John Wiley and Sons.

Bourgois, P. (1996). In search of masculinity violence, respect and sexuality among Puerto Rican crack dealers in East Harlem. *British Journal of Criminology, 36*, 412–427.

Bourgois, P. (2003). *In search of respect: Selling crack in El Barrio*. New York, NY: Cambridge University Press.

Brownstein, H. H., Mulcahy, T. M., & Huessy, J. (2014). *The methamphetamine industry in America: Transnational cartels and local entrepreneurs*. New Brunswick, NJ: Rutgers University Press.

Brownstein, H. H. (2013). *Contemporary drug policy*. Oxford, UK: Routledge Books.

Brownstein, H. H. (2000). Drug distribution and sales as a work system. In C. Faupel & P. M. Roman (Eds.), *Encyclopedia of criminology and deviant behavior—Volume four, self-destructive behavior and disvalued identity*. London, UK: Taylor and Francis.

Brownstein, H. H., Spunt, B. J., Crimmins, S., & Langley, S. (1995). Women who kill in drug market situations. *Justice Quarterly, 12,* 473–498.

Brownstein, H. H., Baxi, H. R. S., Goldstein, P. J., & Ryan, P. J. (1992). The relationship of drugs, drug trafficking, and drug traffickers to homicide. *Journal of Crime and Justice, 15,* 25–44.

Brownstein, H. H. (1990). Surviving as a qualitative sociologist: Recollections from the diary of a state worker. *Qualitative Sociology, 13,* 149–167.

Brownstein, H. H. (1983). The adequacy of intensive interview data: Preliminary suggestions for the measurement of validity. *Humanity and Society, 7,* 301–320.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Campbell, H. (2005). Drug trafficking stories: Everyday forms of narco-forklore on the U.S.-Mexico border. *International Journal of Drug Issues, 16,* 326–333.

Cherryholmes, C. H. (1988). Construct validity and the discourses of research. *American Journal of Education, 96,* 421–457.

Creswell, J. W. (2003). *Research design—Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Crimmins, S. M., Cleary, S., Brownstein, H. H., Spunt, B., & Warley, R. (2000). Trauma, violence and drugs among juvenile offenders. *Journal of Psychoactive Drugs, 32,* 43–54.

Crimmins, S. M., Langley, S., Brownstein, H. H., & Spunt, B. (1997). Convicted women who have killed children: A self psychology perspective. *Journal of Interpersonal Violence, 12,* 49–69.

Denzin, N. K., & Lincoln, Y. S. (1994). Introduction—The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications.

De Quincey, T., & Morrison, R. (2013). *Confessions of an English Opium-Eater and other writings*. New York, NY: Oxford University Press.

Dunlap, E., & Golub, A. (2011). Drug markets during the Katrina disaster. *Disaster Prevention Magazine, 20,* 251–265.

Federal Bureau of Investigation. (2014). *Crime in the United States, 2013*. Washington, DC: U.S. Department of Justice.

Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology, 40,* 285–327.

Gitlin, T. (1993). *The sixties—Years of hope, days of rage* (Revised edition). New York, NY: Bantam Books.

Glaser, B. G., & Strauss, A. L. (2012/1967). *The discovery of grounded theory—Strategies for qualitative research*. Chicago, IL: Aldine-Atherton.

Goldstein, P. J. (1985). The drugs/violence nexus: A tripartite conceptual framework. *Journal of Drug Issues, 15,* 493–506.

Goldstein, P. J., Brownstein, H. H., Ryan, P. J., & Bellucci, P. A. (1989). Crack and homicide in New York City: A conceptually-based event analysis. *Contemporary Drug Problems, 4,* 651–687.

Goldstein, P. J., Spunt, B. J., Miller, T., & Bellucci, P. (1990). Ethnographic field stations. In E. Lambert (Ed.), *The collection and interpretation of data from hidden populations*. NIDA Research Monograph 98. Rockville, MD: National Institute on Drug Abuse.

Goldstein, P. J., Brownstein, H. H., & Ryan, P. J. (1992). Drug-related homicide in New York: 1984 and 1988. *Crime and Delinquency, 38,* 459–476.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method*. New York, NY: Oxford University Press.

Gusfield, J. R. (1986). *Symbolic crusade: Status politics and the American temperance movement*. Urbana, IL: University of Illinois Press.

Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in practice*. London, UK: Tavistock Publications.

Heckathorn, D. D. (2007). Extensions of respondent-driven sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociological Methodology, 37,* 151–207.

Heckathorn, D. D. (1997). Respondent driven sampling: A new approach to the study of hidden populations. *Social Problems, 44,* 174–199.

Hesse-Biber, S. N., & Leavy, P. (Eds.). (2004). *Approaches to qualitative research—A reader on theory and practice*. New York, NY: Oxford University Press.

Hughes, P. (1977). *Behind the wall of respect: Community experiments in heroin addiction control*. Chicago, IL: University of Chicago Press.

Hunt, G., Joe, K., & Waldorf, D. (1997). Culture and ethnic identity among Southeast Asian Gang members. *Free Inquiry, 25,* 9–21.

Inciardi, J. A. (1992). *The war on drugs II: The continuing epic of heroin, cocaine, crack, crime, AIDS, and public policy*. Mountain View, CA: McGraw-Hill.

Jaynes, G., Apter, D., Gans, H. J., Kornblum, W., Horowitz, R., Short, J. F., et al. (2009). The Chicago school and the roots of urban ethnography. *Ethnography, 10,* 375–396.

Johnson, B. D., Goldstein, P. G., Preble, E., Schmeidler, J., Lipton, D. S., Spunt, B., et al. (1985). *Taking care of business—The economics of crime by heroin abusers*. Lexington, MA: Lexington Books.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. San Francisco, CA: Chandler.

Koester, S. K. (1994). Copping, running, and paraphernalia laws: Contextual variables and needle risk behavior among injection drug users in Denver. *Human Organization, 53,* 287–295.

Lambert, E. Y. (1990). *The collection and interpretation of data from hidden populations*. NIDA Research Monographs. Document Publication Number (ADM) 90-1678. Rockville, MD: National Institute on Drug Abuse.

Lambert, E. Y., & Wiebel, W. W. (1990). Introduction. In E. Y. Lambert (Ed.), *The collection and interpretation of data from hidden populations*. NIDA Research Monographs. Document Publication Number (ADM) 90-1678. Rockville, MD: National Institute on Drug Abuse.

Lazarsfeld, P. F., & Rosenberg, M. (Eds.). (1955). *The language of social research: A reader in the methodology of social research*. New York, NY: Free Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage.

Lindesmith, A. R. (1947). *Opiate addiction*. Oxford, UK: Principia Press.

Lindesmith, A. R. (1938). A sociological theory of drug addiction. *American Journal of Sociology, 43,* 593–613.

Maher, L. (2000). *Sexed work: Gender, race and resistance in a Brooklyn drug market*. New York, NY: Oxford University Press.

Maloney, M., MacKenzie, K., Hunt, G., & Joe-Laidler, K. (2009). The path and promise of fatherhood for gang members. *British Journal of Criminology, 49,* 305–325.

Martin, A., & Stenner, P. (2004). Talking about drug use: What are we (and our participants) doing in qualitative research? *The International Journal of Drug Policy, 15,* 395–405.

McElrath, K., & O'Neill, C. (2011). Experiences with mephedrone pre-and post-legislative controls: Perceptions of safety and sources of supply. *International Journal of Drug Policy, 22,* 120–127.

Messick, S. (1975). The standard problem—Meaning and values in measurement and evaluation. *American Psychologist, 30,* 955–966.

Murphy, S., & Rosenbaum, M. (1999). *Pregnant women on drugs: Combating stereotypes and stigma*. New Brunswick, NJ: Rutgers University Press.

Murphy, S., Sales, P., & Averill, S. (2016). Ethnographic studies of drugs in communities. In H. H. Brownstein (Ed.), *The Handbook of Drugs and Society*. Hoboken, NJ: John Wiley and Sons.

National Drug Intelligence Center. (2014). *National drug threat assessment 2014*. DEA-DCT-DIR-002-15. Washington, DC: US Department of Justice.

National Institute of Justice. (2003). *Annual report 2000—Arrestee drug abuse monitoring*. Washington, DC: National Institute of Justice, US Department of Justice.

Office of National Drug Control Policy. (2014). *ADAM II. 2013 annual report. Arrestee drug abuse monitoring program II.* Washington, DC: Office of National Drug Control Policy.

Preble, E., & Casey, J. J. (1969). Taking care of business—The heroin user's life on the street. *International Journal of the Addictions, 4,* 1–24.

Ragin, C. C., & Amoroso, L. M. (2011). *Constructing social research—The unity and diversity of method* (2nd ed.). Newbury Park, CA: Pine Forge Press.

Rasche, C. E. (1990). Early models for contemporary thought on domestic violence and women who kill mate: A review of the literature from 1895 to 1970. *Women and Criminal Justice, 1,* 31–53.

Rhodes, T., & Moore, D. (2001). On the qualitative in drugs research: Part one. *Addiction Research & Theory, 9,* 279–297.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review, 16*, 812–818.

Roszak, T. (1968). *The making of a counter culture*. Oakland, CA: University of California Press.

Ryder, J., Langley, S., & Brownstein, H. H. (2009). I've been around and around and around. Measuring traumatic events in the lives of incarcerated girls. In R. Gido & L. Dalley (Eds.), *Women's mental health issues across the criminal justice system.* Englewood Cliffs, NJ: Prentice Hall.

Sadler, R., Lee, H. C., Lim, R. S. H., & Fullerton, J. (2010). Recruiting hard-to-reach United States population sub-groups via adaptations of snowball sampling strategy. *Nursing Health Science, 12,* 369–374.

Sales, P., & Murphy, S. (2007). San Francisco's free-lancing ecstasy dealers: Towards a sociological understanding of drug markets. *Journal of Drug Issues, 37,* 779–814.

Schutz, A. (1954). Concept and theory formation in the social sciences. *The Journal of Philosophy, 51,* 257–273.

Shaw, C. (1930). *The Jack-Roller: A delinquent boy's own story*. Chicago, IL: University of Chicago Press.

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology, 37,* 57–86.

Spunt, B. J., Goldstein, P. J., Brownstein, H. H., Fendrich, M., & Langley, S. (1994). Alcohol and homicide: Interviews with prison inmates. *Journal of Drug Issues, 24,* 143–163.

Spunt, B. J., Brownstein, H. H., Crimmins, S., & Langley, S. (1996). American women who kill: Self-reports of their homicides. *International Journal of Risk, Security, and Crime Prevention, 1,* 293–303.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2012). *Treatment episode data*

set (TEDS): 2010–11. National admissions to substance abuse treatment services. DASIS Series S-61, HHS Publication No. (SMA) 12-4701, Rockville, MD: SAMHSA.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2013). *Highlights of the 2011 the DAWN Report: Highlights of the 2011 drug abuse warning network (DAWN) findings on drug-related emergency department visits*. Rockville, MD: SAMHSA.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2014). *Results from the 2013 national survey on drug use and health: Summary of national findings*. NSDUH Series H-48, HHS Publication No. (SMA) 14-4863. Rockville, MD: SAMHSA.

Sutherland, E. H. (1939). *Principles of criminology*. New York, NY: JB Lippincott.

Thoumi, F. E. (2005). The numbers game: Let's all guess the size of the illegal drug industry! *Journal of Drug Issues, 35,* 185–200.

Thrasher, F. M. (1927). *The Gang*. Chicago, IL: University of Chicago Press.

Turner, R. H. (1953). The quest for universals in sociological research. *American Sociological Review, 18*, 604–611.

Turner, D. W., III. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report, 15,* 754–760.

van Meter, K. M. (1990). Methodological and design issues: Techniques for assessing the representatives of snowball samples. In E. Y. Lambert (Ed.), *The collection and interpretation of data from hidden populations*. NIDA Research Monographs. Document Publication Number (ADM) 90-1678. Rockville, MD: National Institute on Drug Abuse.

Venkatesh, S. (2008). *Gang leader for a day—A rogue sociologist takes to the streets*. New York, NY: Penguin Press.

Waldorf, D. (1973). *Careers in dope*. Englewood Cliffs, NJ: Prentice Hall.

Waldorf, D., Reinarman, C., & Murphy, S. (1992). *Cocaine changes: The experience of using and quitting*. Philadelphia: Temple University Press.

Warren, C. A. B., & Karner, T. Z. (2010). *Discovering qualitative methods—Field research, interviews, and analysis* (2nd ed.). New York, NY: Oxford University Press.

Weber, M. (1947). *The theory of social and economic organization* (Ed. and tr. by A. M. Henderson and T. Parsons). New York, NY: Free Press.

Wiebel, W. W. (1990). Identifying and gaining access to hidden populations. In E. Y. Lambert (Ed.), *The collection and interpretation of data from hidden populations*. NIDA Research Monographs. Document Publication Number (ADM) 90-1678. Rockville, MD: National Institute on Drug Abuse.

Williams, T. M. (1992). *Crackhouse: Notes from the end of the line*. Reading, MA: Addison-Wesley Publishing Company.

Znaniecki, F. (1934). *The method of sociology*. New York, NY: Farrar & Rinehart.

Zorbaugh, H. W. (1929). *The gold coast and the slum*. Chicago, IL: University of Chicago Press.

Using GIS for Substance Abuse Research and Intervention

9

Jacqueline W. Curtis and Andrew Curtis

9.1 Background

From the iconic visualization of cholera cases and the Broad Street pump by Dr. John Snow in nineteenth century London to the emergent applications of geographically enabled biosensors, maps, and the geospatial techniques used to create them are essential tools to understand the health–place nexus. The main objective of this chapter is to raise awareness of the applications of mapping in public health with regard to issues related to substance abuse. However, we begin with an overview of Geographic Information Systems (GIS), which is the primary tool used to create maps, and allied geospatial technologies (GT) (e.g., Global Positioning System (GPS), spatial analytic software). The remainder of the chapter is devoted to reviewing the extant research on substance abuse that utilizes these methods. We then move to a discussion of "what's next," in essence the emergent approaches that are being adopted to further advance scholarship on a myriad of issues related to understanding the relationship between health and place, with a focus on substance abuse research.[1]

9.2 A Brief Primer on Geographic Information Systems (GIS) and Geospatial Technology (GT)

Maps have long been a central tool to document and disseminate observations, enabling amalgamation of these observations to provide synthesis in understanding places and patterns over space. However, the advent of GIS and the concomitant technical revolution of the 1960s to present have meant that more data can be analyzed in more ways by more people. This situation has led to the diffusion of GIS and spurred advances in research in the earth and social sciences, as well as the humanities (e.g., HGIS). Today, GIS has a number of slightly different definitions, but most agree that it is the combination of hardware and software that enables visualization (e.g., mapping) and analysis of spatial data. Spatial data are any phenomena that have a location, ranging from more coarse scales such as occurring globally (drug trafficking routes), through regional (overdose rates by state or city) to finer

J.W. Curtis (✉) · A. Curtis
GIS Health & Hazards Laboratory, Department of Geography, Kent State University, 325. S. Lincoln St, Kent, OH 44242-0001, USA
e-mail: jmills30@kent.edu

A. Curtis
e-mail: acurti13@kent.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_9

[1]The call to understand context is not new in substance abuse research (Dembo et al. 1985), however, systematically collecting and analyzing contextual data in a way that is replicable and extensible has been difficult. It is in these areas that GIS and other geospatial technologies contribute to contextual understanding.

spatial scales such as a zip code (drug arrests throughout Chicago), or street address (a neighborhood clinic providing drug counseling). GIS operates with dynamically linked windows that usually include a database and a map. Each row in the database represents a unit in the map (e.g., a point representing a patient's residential address); each column in the database is a characteristic of that unit (e.g., the age of the patient, the source of the addiction). Once these data are in a GIS, there are a range of ways to map and spatially analyze them in order to understand patterns (e.g., clusters of discarded needles) and relationships (e.g., what environments are around the needles as identified by spatial regression).[2] The resulting visualizations can then be used as exploratory tools to generate hypotheses or simply to communicate situation awareness within research teams, to guide intervention policy, or if appropriate, to disseminate to the public. Depending on the form of analysis, the map may even provide answers to practical, time-sensitive questions such as where to send a mobile health unit.

Geospatial techniques encompass a range of approaches for data collection, analysis, and visualization that are linked to location, with GIS the platform through which all of these data are organized and managed, analyzed for relationships across space and over time, and visualized in the form of a map. In substance abuse research, census units (e.g., zip code tabulation areas (ZCTAs), tracts, blocks), health service regions, and addresses have traditionally served as the geographic unit of analysis. However, with GPS technology, even the moment-to-moment movements of an individual entity (e.g., person, product) can be mapped and analyzed in the GIS. Traditionally, GIS has been utilized to map and analyze objective or official spatial data, such as demographic data or the location of certain assets or hazards (e.g., hospitals, parks, bars, crime incidents). More recently, GIS and allied geospatial techniques have been employed to understand and integrate "new" data, such as the locations of needles, or qualitative data, such as individual perceptions and/or local knowledge of place. This paradigmatic shift is revealing more than understanding spatial patterns of phenomena or outcomes (*where?*), but also the context, mechanisms, and explanations for these patterns (*why there?*).

Despite the power of GIS, they are only as good as the available data and the relevant knowledge of the person using the technology. Many texts and courses offer the training needed to be a capable GIS user and we strongly suggest taking advantage of these resources.[3] However, for the purpose of this chapter in providing a brief overview, we note some common considerations in use: choosing a data model, data input methods, spatial scales, forms of spatial analysis, and map dissemination. All of these issues should be considered prior to beginning a GIS project and should be guided by the overarching purpose of the project and the needs of the stakeholders. We will first address each of these issues in general while placing them in the context of substance abuse-related examples. In a subsequent section we will revisit them as they relate to the existing literature on substance abuse research and practice.

9.2.1 Data Models

GIS data fall into in one of two models, vector or raster, with vector data the most commonly used form in public health and across the social sciences. The vector model represents spatial data as discrete objects: points, lines, and polygons. For example, points could be used for treatment

[2]A number of texts provide in-depth discussion of mapping and spatial analysis techniques and issues. See, for example, texts by Dent et al. (2009) and Maguire et al. (2008).

[3]Most universities with a geography department offer GIS courses. However, GIS can also be taught in other departments as well. Check your local university or community college for GIS course availability. In addition, GIS manuals are useful to novices and experienced users alike. Finally, all makers and users of maps are advised to read Monmonier's (1996) classic text, "How to Lie with Maps" for insight on the use and abuse of maps, including how to evaluate maps critically.

centers, lines for roads, and polygons for a clinic's service area. Figure 9.1 shows methadone clinics as points and 1.5 km accessible zones as well as government boundaries as polygons. Alternatively, the raster model represents data as continuous, a series of grid cells coded based on a characteristic such as temperature or elevation. Raster data are more commonly used in the earth sciences, but can certainly be integrated with vector data for health investigations. Despite the many benefits of GIS, forcing data into points, lines, polygons, or grid cells can be a limitation, as they are not always optimal or even appropriate for certain types of data, especially those where there is some uncertainty about location or some "fuzziness" in boundaries (e.g., walkable distances, the activity space of the homeless). Data model selection is important as it has implications in terms of how data are collected, and then spatially analyzed.

Throughout this chapter, to place "data models" in the context of substance abuse, we are using an example of studying teen drug overdose. In this case, we would probably include several spatial data layers that include the locations of drug overdoses, along with locations of clinics and hospitals, and the address of the residence. These data would be represented in a GIS using a vector data model, with all of these locations as point data. If we were investigating how the physical environment is predictive of the likelihood of overdoses, then we might use raster data to assess the percentage of green space to buildings.

9.2.2 Data Input

Data input is achieved primarily through the following mechanisms: downloading an existing GIS file from a secondary source (for example, demographic data from the census), geocoding (meaning placing a known spatial reference on a map, such as a residential street address), digitizing (creating new data by "clicking the mouse to draw" on an underlying map, such as an aerial photograph), and bringing in coordinates from a GPS. Usually, multiple forms of data input are used in a GIS project, such as census tract demographic data with geocoded address data of treatment centers or study participants. Each of these common data input strategies is now addressed, along with considerations for their use.

Secondary sources of spatial data are most commonly government agencies, and even states, counties, or cities. Indeed, many of these entities have an online atlas or other webpage specifically devoted to sharing their data in a GIS format (e.g., shapefile, layer).[4] Even when government data are not available in a GIS format, if they have any locational attributes attached, then they can be added to a GIS using geographic identifiers. Perhaps the most common data that fit this description are U.S. Census tables generated by American Factfinder.[5] In order to input these data to GIS all that is needed is a cartographic boundary file such as census tracts, block groups, or blocks with a geographic identifier in the attribute table. The tabular data from the census should have the same identifier and the two files can be joined based on this common column. This approach is also regularly used when working at the state or county scale to make a map of meth use, for example.

Returning to an example, we might want to locate drug overdoses with the census characteristics of both where the event occurred, and the residence of the patient. Our research questions might then consider how the socioeconomic status of the area predicts the overdose, and even type of overdose (these socioeconomic data can be acquired from the U.S. census). By adding in school districts and educational performance, we again might be able to use these as predictors (these data can be acquired from state departments of education or local school boards).

Another common approach for data input is geocoding. In essence, this procedure uses

[4]Some examples of these sources include: Atlas—The Louisiana Statewide GIS (http://atlas.lsu.edu/), Cal-Atlas (http://www.atlas.ca.gov/download.html), SanGIS (http://www.sangis.org/).

[5]Go to http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml and then use the "Download Center" to access data tables with geographic identifiers such as tract, block group, and block.

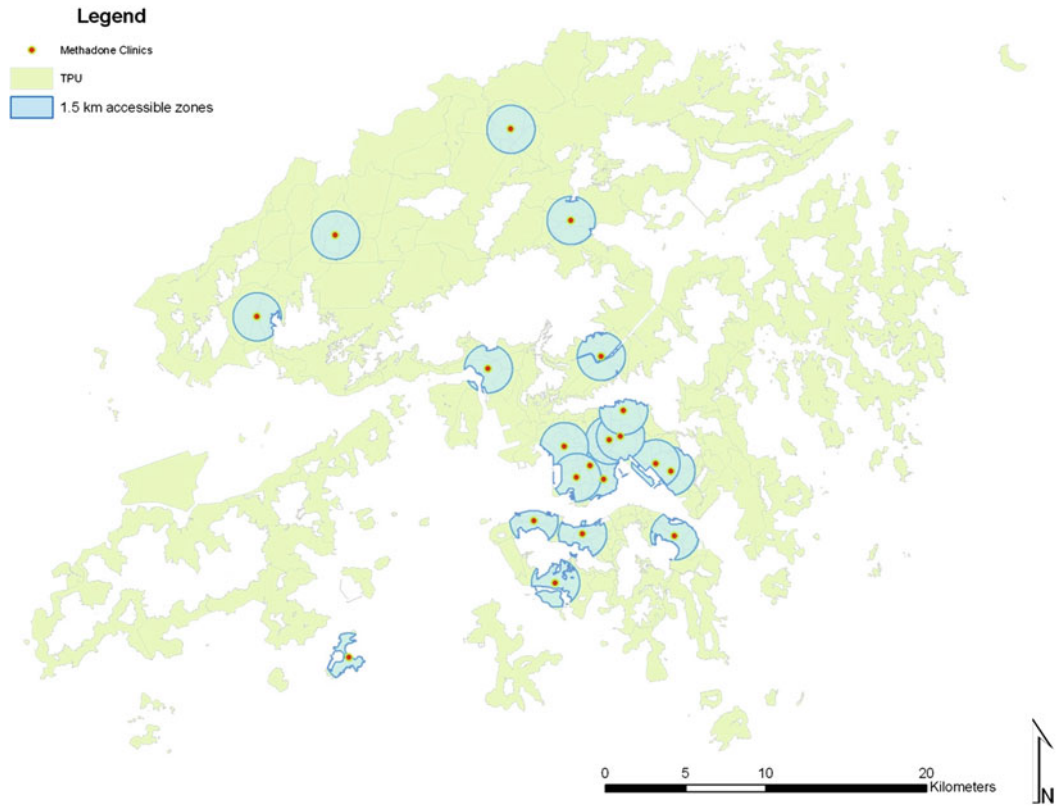

Fig. 9.1 Example of points and polygons in a vector data model. See Pang and Lee (2008) for full text on the model

addresses to assign points to their location on the map. Geocoding is typically used with street address data (e.g., street number, street name, city, state, zip code), which can provide good positional accuracy of the point.[6] However, less accurate geocoding procedures also exist, such as assigning data to zip code or street segment. When using a complete address, geocoding works by placing the street number proportionally along a street segment. For example, if the address range for Main Street is 100–500 on one side of the road, then the address 300 Main Street will be placed at the midpoint of this street segment. Clearly, this approach will lead to better positional accuracy in some areas than in others as addresses are not always placed proportionally in the real world (e.g., suburban vs. rural areas).[7] More areas are now improving the accuracy and precision of these data by using a GPS to get the exact coordinate for each address.

In our example the most obvious use of geocoding is the location of where the overdose occurred, the residence of the individual, and maybe even the location of a clinic providing care to this patient.

The next two approaches for data input in GIS are more time-intensive. First, digitizing is the process of "drawing" data in the GIS. It can be employed to capture features from aerial imagery (e.g., building footprints, sidewalks) or from paper maps (e.g., participants' perceived areas where drug crime is highest).[8] Digitizing can be a

[6]Positional accuracy refers to the degree to which a feature's location on the map matches its real-world location.

[7]See, for example, article by Mazumdar et al. (2008).

[8]A growing body of research is emerging on this approach to understand the environment–behavior nexus. See the following sources for a complete list of references and discussions of the prospects and problems of integrating sketch maps with geographic information systems to understand environmental perception: Curtis (2012), Curtis et al. (2014), and Curtis (2016).

powerful way to create new spatial data for integration in maps and GIS analysis. However, digitizing features can be time-intensive, depending on the scale of the study and the nature of the features being digitized. In recent years as aerial imagery is being provided in geographically enabled formats (e.g., the images have coordinates so that a GIS can place them in their real-world location), digitizing from these sources is becoming easier, especially through online geospatial software such as Google Earth. However, digitizing from paper maps or from sources that are not geographically enabled still requires the first step of georegistering the image. This means that the GIS user has to insert *x*, *y* coordinates into the graphic so that the GIS knows where to place it. Only after this step is successfully accomplished can digitizing occur.

In our example, digitizing would be an appropriate form of data input if we are interested in asking teenagers where they go for drugs, and where they then use the drugs; their answers may come in the form of sketch maps. These can be georegistered, and the locations digitized in the GIS to form a previously unavailable map of "usage."

Finally, GPS receivers can be used to collect data in the field. This technology can collect points and paths which can then be imported into the GIS. Like digitizing, this is a more time-intensive approach. Furthermore, it is important that the user has a basic understanding of how GPS works, when to collect data (e.g., mission planning), and the errors and accuracy issues with this form of technology.[9] Points and paths are collected in the field and are then uploaded into GIS. There are a few ways to accomplish this task (including bringing the data in first through Google Earth), but in essence, each point collected has a latitude and longitude value. With these coordinates, the GIS can bring in the data points and place them at the real-world location. For example, the blue points displayed in Fig. 9.2 represent the path driven in a car while conducting a built environment survey in one neighborhood of Akron, Ohio. GPS receivers can be set to collect a *x*, *y* location at set time intervals, in this case every 1 s. This is but one of a growing number of examples of GPS-enabled data collection. As GPS receivers have become smaller, easier to use, and less expensive, it has received increasing use across the social sciences and public health. In particular, this geospatial technology offers substance abuse researchers the potential to systematically collect new data on the activity patterns of study participants, as well as observations of the natural, built, and social environments that would otherwise not exist. GPS is particularly important in collecting data that are dynamic or ephemeral (e.g., human behavior in place).

In our example the GPS coordinates along the route could link to the "geonarrative" of a teenager riding in the car. He/she could describe the neighborhood and locations where drugs are bought and where they are used, which would then be mapped by linking the transcribed word to its coordinate.

Clearly, there are a number of ways to bring data into a GIS. It is important to note that making the decision on data input may not be within your control as it is frequently the result of what existing data are available to you and/or cost constraints associated with a study. This situation may also dictate the spatial scale of the mapping and spatial analysis. However, it is wise to have an idea of the variation of spatial scales at which GIS can operate and the influence these decisions have on the benefits and limitations of your resulting maps.

9.2.3 Spatial Scales

In Fig. 9.2, data are collected and displayed at a fine spatial scale (individual *x*, *y* coordinates and by parcel). Such scales enable investigations of microgeographic patterns and processes, such as the drug activity and health outcomes along a single street. With the expanding use of GPS and GPS-enabled sensors, this scale is receiving

[9]Many websites and texts are devoted to helping users understand GPS and use it appropriately. The U.S. government has a good overview of the technological aspects of GPS: http://www.gps.gov/.

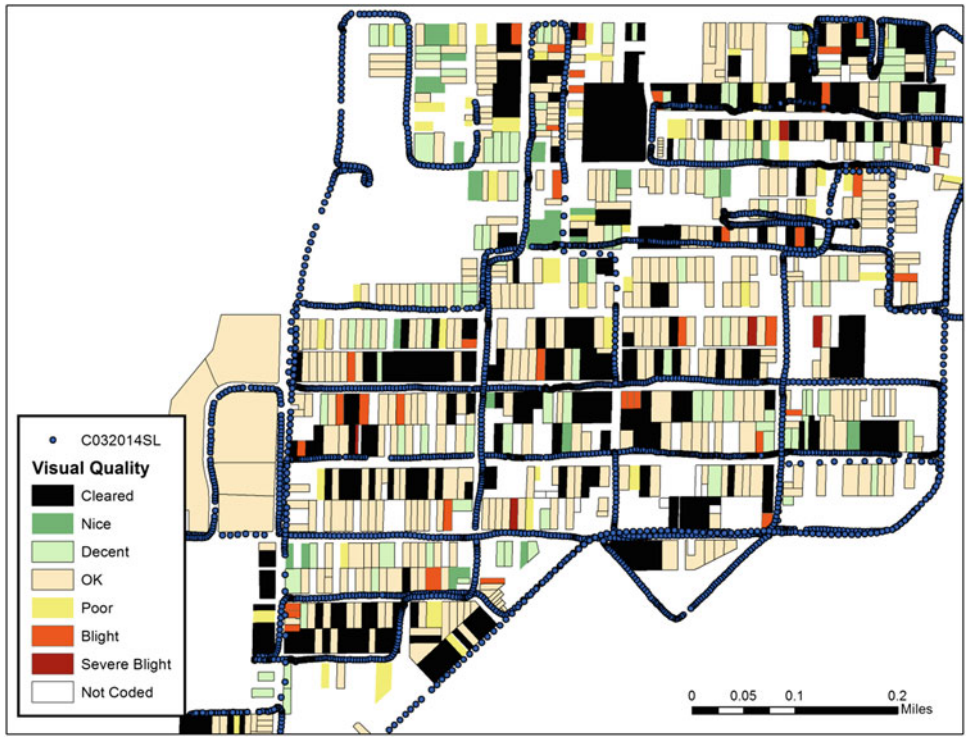

Fig. 9.2 Example of GPS track that shows the path driven for a built environment survey of parcels within a neighborhood in Akron, Ohio

increasing attention as the technology now exists to work systematically on small area analysis. However, a limitation of this work is that it is time-intensive as it requires the need to conduct primary field-based data collection. More commonly, GIS is utilized to work at coarser scales and using government-defined boundaries (e.g., census tracts). This makes sense as data are easily accessible, are ubiquitous and comparable across the country, and contain unmatched demographic data for inclusion in analysis. The downside to this coarser scale is that it often misses the context of day-to-day living (e.g., many of us are not at home all day, though census data are linked to residential location).

Another problem when working with such aggregated space is the potential for the ecological fallacy and the Modifiable Areal Unit Problem (MAUP).[10] The ecological fallacy is based on the premise that characteristics for the unit, such as there being 30% of properties being blighted in a census tract, are not representative of characteristics of any one individual in that unit; it would be false to assume that for any randomly picked street 30% of properties are blighted. Although U.S. Census units are designed based on some degree of homogeneity, there can still exist considerable spatial heterogeneity. Returning to the previous example, it is likely that certain streets within that census tract are all blighted, because this type of urban decay clusters. Another issue related to geographic scale is the MAUP, which has two effects: zone and scale. The zone effect refers to the boundary of the area for analysis. Although a census tract and a zip code are approximately the same physical size (for arguments sake), if the boundaries are slightly different, then the overall value for that area may change if a high intensity of events falls just inside or outside. This is certainly possible as roads are often the

[10]For a complete discussion of these issues, see Longley and Batty (1996) and Openshaw (1996).

boundaries of both units. Remember that blight cluster? It might fall just inside the zip code, but outside the Census Tract. The colored map of blight will then look different for what is approximately the same area.

The scale effect is the size of the area being analyzed, which for the U.S. Census include ZCTAs, tracts, block groups, or blocks. Figure 9.3 displays a hot spot of drug sales and violations in Akron, Ohio with these four administrative units. The hot spot was created using geocoded address data which does not adhere to any preset spatial unit. You can see that this hot spot does not perfectly correlate with any aggregated unit. Even in the largest unit, the ZCTA, the hot spot still slightly crosses over into two adjacent boundaries. If we were to make a map of drug sales by any one of these aggregations, then the area and size of the intensity would change. A map-reader might then conclude that "this census tract needs intervention," where the reality is it is a few streets, or a block, that need these targeted resources. There is a lot of the geography within that census tract, which is now colored "problematic" that is actually good. If at all possible is it worth mapping data at different scales just to get a more realistic impression of the problem.

This variety of geographic scales to map and analyze data presents both benefits and limitations. Primary considerations for selecting the zone and scale of analysis should be the purpose of the research and the needs of the stakeholders. Use of spatial aggregation units may be appropriate as a compromise to show important patterns to policy makers and the public, while preserving the spatial confidentiality of individuals. As alluded to earlier, these predefined government units may be inappropriate to represent actual human use of geographic space and so maps and analytical results should be interpreted with care.[11]

In our example, we might make a map of teenage drug overdoses by county within Ohio, and this will show us where the problems exist regionally. If we change scale to the census tract, we can focus in on those problem areas to show where within the county overdoses occur. In reality, most overdoses are likely to be within tight geographic spaces, around parks, or street corners, or in certain abandoned houses—so our maps need to get finer still, down to the block groups and even blocks to really show the environments of drug overdoses.

The issues raised in this section should not prevent use of predefined scales, but prior to project commencement, time should be spent on investigating appropriate scales of data collection and analysis in order to most appropriately answer the research question and meet the needs of stakeholders.

9.2.4 Spatial Analysis

Once scales have been determined and data collected, spatial analysis enables making sense of the resulting maps. Ideally, determining the most appropriate forms of analysis has occurred prior to project implementation as the type of data (raster or vector; points, line or polygon) and the scale of investigation will have some bearing on the methods of analysis that are available. Spatial analysis can take many forms, though the most common approaches involve searching for hot spots and modeling relationships. In addition, the power of visual analysis (just looking at a map) is also an important tool. There can be great power in mapping data in a variety of ways and observing patterns and relationships; this can be helpful for exploratory analysis and hypothesis generation as well as a way to simply show spatial distributions and proposed relationships by examining how different data are spatially related to each other through overlay (meaning stacking several maps one on top of the other). Again, as vector data (points, lines, polygons) are more common than raster data use in public health, we will focus on general forms of spatial analysis appropriate for this format. Points and

[11]For a more in-depth discussion of these issues, see Coulton et al. (2001) or Curtis et al. (2015a, b).

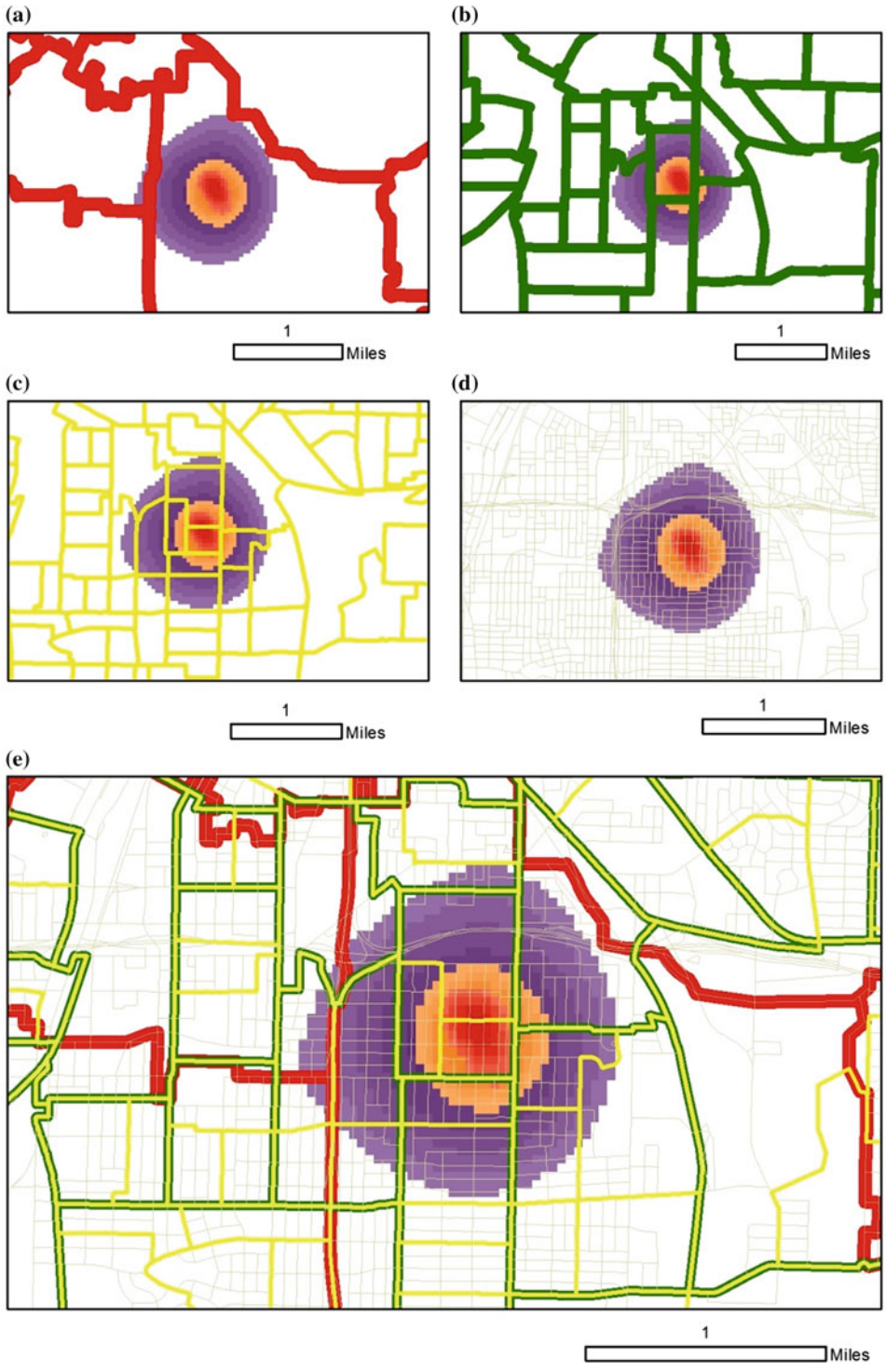

Fig. 9.3 Example of varying geographic scales of spatial data in GIS: **a** ZCTA, **b** tract, **c** block group, **d** block, **e** all scales together. Hot spot: based on police Calls for Service data for Drug Sales/Violations at the address level (2014)

polygons are the most common data used in spatial analysis, with line data less so.[12]

Returning to the earlier overview of data models, points are traditionally used to represent a variety of health outcomes, most commonly the residential address of a subject. They are also used to represent the locations of certain health-related events, such as violent criminal acts, or health-related locations such as treatment centers, hospitals, and physicians' offices. When these data are geocoded, the result is a map covered in points, through which it can be difficult to see patterns or investigate spatial relationships. For this reason, one of the first spatial analytical approaches used is often Kernel Density Estimation (KDE) which identifies areas of elevated concentration of points based on a user-defined search window or "kernel."[13] Density of points within the kernel is determined where locations closer to the center have more weight in the calculation than those further away, up until the boundary of the kernel. Returning to Fig. 9.3, The hot spot of drug sales/violations based on police data was created with KDE at three different kernel sizes (also called neighborhood or bandwidth): 400 m = red, 800 m = orange, 1600 m = purple. This approach offers a visualization of point density, but it is important to note, as with the previous discussion about scale, that the hot spot can vary based on the bandwidth selected by the user. Though there is some widely acknowledged subjectivity in this approach, it nonetheless is an accepted way to identify hot spots. For those who have the requisite data, skill and need, statistically rigorous cluster identification techniques and approaches for space-time cluster detection are also available.[14]

In addition to spatial analytical approaches for point data, several techniques have also been developed to identify hot spots in polygon data. Again, polygons are most commonly census aggregation units (ZCTAs, tracts, block groups, and blocks) with the focus of analysis usually being demographic data. While the type of granular hot spot identification is not as possible as it was with points, more coarse areas of intensity, as well as cold spots and outliers can be identified through approaches such as Anselin's Local Moran's I (a part of Local Indicators of Spatial Autocorrelation—LISA) and Getis and Ord's Gi*.[15] Finally, polygon and point data can also be used as the unit of analysis in a spatial regression, such as Geographically Weighted Regression (GWR).[16] Unlike existing modeling approaches, GWR enables coefficients to be geographically varied which results in local models of spatial relationships, rather than only one global model. In essence, GWR explicitly takes into account the spatial nature of processes that lead to certain outcomes, which are not necessarily the same in all places. Despite its widespread use across the sciences, GWR has received comparatively little use in substance abuse investigations.

In our example, we want to make sense of where those teen drug overdoses occur. Initially, and in an exploratory mode, we find where the hotspots are located. These hotspots might be of actual overdoses, and depending on the scale of the analysis, we might be able to identify key locations, such as parks, corners, or houses. If we include the type of housing found in Fig. 2, we could then run a GWR to predict how the proportion of blighted properties around a school is predictive of the number of teen drug overdoses.

The output of this framework of GIS project design and implementation (models, input, scale, analysis), is the map. Ostensibly, it can show where problems are located, their social–environmental context, and even probable causes for

[12]Through advances in tools such as ArcGIS Network Analyst (ArcGIS 10.2. ESRI. Redlands, CA), line data are increasingly being analyzed in public health research, for example with ambulance response times, or even travel patterns of the homeless.

[13]For a more detailed discussion of KDE for public health applications, see Carlos et al. (2010).

[14]Spatial filtering, SpaceStat; In addition, space-time clusters can be identified through SatScan; see Fotheringham (1997).

[15]See Anselin (1995) and/or Getis and Ord (1992).

[16]See Fotheringham et al. (2002).

the resulting geographic pattern. However, the map itself must be addressed in terms of what it looks like and how it is utilized. This issue is particularly important when dealing with sensitive health data. Given its importance, we conclude the overview with a discussion of map dissemination.

9.2.5 Map Dissemination

With an understanding of GIS data models, the scales of their use, methods of data input, and approaches for spatial analysis, we turn to arguably the most important output: the map. It is important to consider who will be looking at the map and how they will be looking at it. First, a decision must be made about whether the results will be private (e.g., only viewable by internal employees or by certain staff) or will they be public. This decision has legal and ethical implications as spatial analysis and mapping of confidential health data may not be able to be released in any form or must go through masking approaches in order to make it suitable for public viewing.[17] If the maps are intended for public use, then it is imperative to consider the following. Ask whether you are adhering to Health Insurance Portability and Accountability Act (HIPAA) guidelines? If no, then clearly no publicly viewable maps can be produced. However, even if you are in compliance, ask yourself the following: Am I displaying health data with other map reference layers (e.g., roads, buildings) which can reveal the location of a person or group of people? It is surprisingly easy to re-engineer locations from published maps and takes only basic GIS skills to do so.[18] However, if the data are not sensitive or if they are adequately masked to preserve individual locations, then it is equally important to make sure that the map is effective and meets the purpose of the project and the needs of the stakeholders. These considerations are addressed through the "cartographic process" or the "cartographic design process"[19] whereby many decisions are made about the maps, including a determination about the format (e.g., interactive webmap, PDF, hard copy paper map, PDF, GeoPDF). Many formats for maps now exist, with some more user-friendly than others.[20] In addition, Google Earth is now more commonly being utilized as a format for map and spatial data dissemination among researchers and with the public.[21]

In our example, a map made for the local newspaper to convey the problem of teenage drug overdoses is colored using gradations of yellow to red at the census tract level. The message gets across without violating any confidentiality. However, in the local hospital, fine scale spatial analysis, overlaid with actual drug overdose locations, is shown to an "intervention" team comprising of social workers and educators. These are on the "inside", can look at sensitive data, and need to know exactly where the problem areas are located. This map is only for their viewing—either in the hospital, or on a mobile device as they go into the field.

With this brief overview of GIS, we now turn to examine how this technology has been utilized in substance abuse research.

9.3 Review of the Extant Research on Geospatial Approaches to Understand Substance Abuse

Literature searches were conducted using three research databases: Google Scholar, Web of Science, and PubMed (MEDLINE) for peer-reviewed, English language articles that have been published on the use of GIS in substance abuse research.[22] Methods sections were

[17]See Boulos et al. (2009) for a more detailed discussion.

[18]See Curtis et al. (2006).

[19]Please see the following link for an overview of this process: https://www.e-education.psu.edu/geog160/node/1882.

[20]See Mills (2009).

[21]See Curtis et al. (2012). Google Earth, in particular, is increasing the accessibility of map use for decision-making across the sciences and by the public.

[22]The search was conducted in June and July 2016. These databases were selected based on their representation of the most general (Google Scholar) to the most specific

then read for any articles where use of GIS or allied geospatial technology was unclear based on title and abstract review. The references section of each article was also read to identify any additional sources that would meet inclusion criteria. Based on this approach, 96 articles met the guidelines.[23] Figure 9.4 displays the trend of published research in this area, which mirrors adoption of GIS in public health, though with a slightly later start date than in some other specialty areas with a tradition of spatial epidemiology. Despite this small delay, GIS-based research in substance abuse seems to be on the assent since 2013. Not only is there an increase in the number of studies, but changes are also occurring in what is being investigated and how these projects are being approached from a geospatial perspective.

Based on content in the abstract and in the methods sections, these articles were coded based on (a) specific substance abuse topic (e.g., what was mapped) and (b) geospatial methodological approach (how it was mapped) (Fig. 9.5).[24]

Based on this review, six topical themes emerged: alcohol outlets/promotion/use, tobacco outlets/promotion/use, drug markets/use, then risk behaviors, harm reduction, and treatment (Fig. 9.5). Of course, there is some overlap in these studies and certainly the themes can be subdivided into more specific veins of research, but these six categories are representative of substantive areas within this body of work. The most prominent topics focused on alcohol outlets and drug markets, including the location of and their relationships to a variety of health outcomes and behaviors. These are followed by work on access to and outcomes from treatment and harm reduction as they relate to access and to the environmental context in which participants live. Apparently less represented in this research are risk behaviors and tobacco. Ascertaining the geographic patterns and spatial relationships of risk behaviors is inherently difficult as they are not easily visible and capturing these data rely on participant insight. The nature of these data could explain the dearth of research in this area. The low representation of tobacco issues studied with GIS might also be explained by the lack of any common spatial data layer.

Within the 96 articles, several did not conform to any of the themes, but are nonetheless worthy of discussion as outliers that differentially contribute to the substance abuse field. Muth and colleagues (2000) used residential locations of participants to examine demographic bias in these study participants. Messiah et al. (2003) explored the potential of modeling the locations of future heroine epidemics. Furthermore, a considerable body of work from Michael Mason uses ecological interview and mapping of individual activity spaces to identify the environments surrounding youth and its impact on their substance use (Mason et al. 2004a, b, 2009, Walker et al. 2006). O'Loghlen et al. (2011) investigate active school travel (AST) and through data from school administrators, the issue of drug abuse is raised as a barrier to this form of child physical activity. Two articles covered such a variety of issues, that they could not be appropriately classified in any meaningful way: Caron and colleagues

(Footnote 22 continued)
(PubMed) sources for this subject. Google Scholar search terms were "GIS" and "substance abuse". However, as PubMed and Web of Science enable greater specificity in designing searches, these queries were structured differently from the Google Scholar search. In PubMed, the titles/abstracts of articles were searched for "GIS" or "Geographic Information System*" and "substance abuse". Similarly, the following search terms were used for topic queries in Web of Science: "GIS" or "Geographic Information System*" and "substance abuse".

[23]It should be noted that as in all literature searches, these results are unlikely to be completely exhaustive of the subject due to limitations in the search terms. However, the number of results is appropriate for the subject area and provides a representative set for review.

[24]One area that has received considerable attention over the years, especially from criminologists and planners, is the relationship between alcohol outlets and crime. Note that this chapter only identifies articles of this type where alcohol abuse is investigated in relation to outlet density and crime. In addition, the understanding that geography matters for substance abuse is not new, but articles were not included unless they explicitly use GIS/GT. This inclusion criterion means that studies which use maps, but without explicit mention of GIS/GT in their methods were excluded (e.g., Fortney et al. 1995, Rockwell et al. 1999).

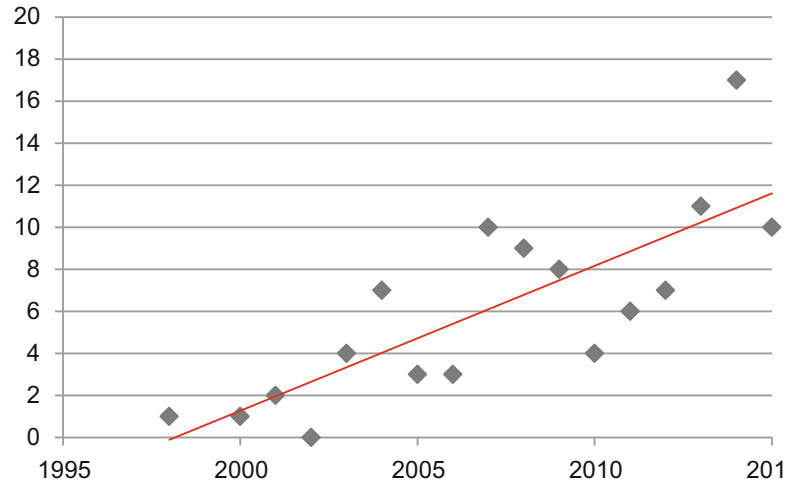

Fig. 9.4 Trend in use of GIS and GT in substance abuse research. This graph reflects results based on the search criteria discussed in this chapter

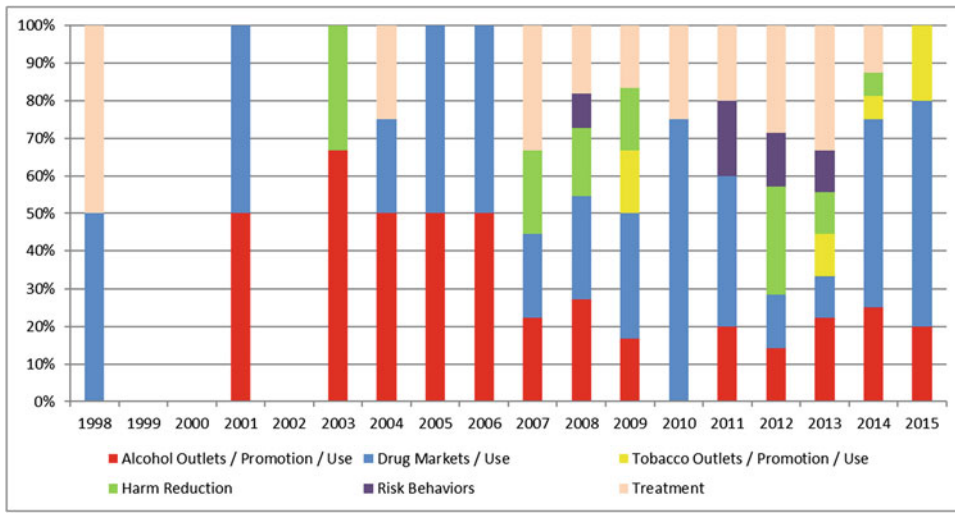

Fig. 9.5 Substance abuse topics under investigation with GIS

(2012) use GIS to "assess the neighborhood social and ecological contexts" (p. 4 of 12), while Milam and colleagues (2013) cover "alcohol, tobacco, and other drugs" (ATOD) from an environmental context. Most recently, Yu et al. (2014) examine the spatial distribution of pharmacies selling addictive products as it relates to socioeconomic status (SES) through use of GWR and Baglivio and colleagues (2015) studied substance abuse only as one of a number of neighborhood level (tract) variables predictive of juvenile delinquent adverse childhood experiences (ACE) scores.

The methods sections of each article were coded for data collection (input) strategies and then for spatial analysis approaches. Four means of data collection were identified: use of secondary data, geocoding, digitizing, and use of GPS. Following the trend observed for topics of investigation with GIS, the methods of data collection and spatial analysis have also diversified with time. Regarding data collection

approaches, geocoding and then use of secondary data are most prominent. Indeed, one fifth of the studies integrated individual level data with other data, usually demographic, from the census aggregated unit in which the point data were located. This most commonly occurred with census tract data. Overall, for the period from 2013 to present alone, a notable increase in diversity of methods is evident with use of digitizing and GPS.

GPS use is surprisingly not identified as a data collection method until 2006 and then not again until 2013 when Tanjasiri and colleagues use a basic handheld GPS receiver, the Garmin etrex, to capture the geographic coordinates of environmental risks and protective features for youth tobacco use in Long Beach, California. Photos and descriptions of these places were provided by youth participants through use of Photovoice. The photos were coded and the GPS waypoints used to calculate average distances from youth participants' home addresses. A logistic regression was used to investigate the relationship between youth tobacco use and exposure to environmental features that expose them to risks or provide protection (Tanjasiri et al. 2013). Concomitantly, Wiehe and colleagues (2013) utilize GPS-enabled cell phones to map the activity paths of female youth participants. These GPS-tracks show a location every 5 s. With such fine-scale spatial and temporal data, the investigators were able to study exposure to criminogenic environments and impact on risk behaviors. This study demonstrates an advance over the existing approach of measuring exposure by placing buffers around a static home or school location, while accounting for the spatial–temporal dynamism of the environment–behavior nexus.

Coding of the methods section for spatial analysis approaches revealed greater diversity over the entire period in which GIS has been used in substance abuse research. The following methods were identified: map analysis, measuring distance, buffers, hot spot/cluster analysis, and spatial regression. Using GIS for distance-based analysis has been accomplished through network analysis (allows for calculation of shortest path, overall distances, and configurations of roads and other transportation systems), use of measurements in aspatial correlation and regression, and through buffers (where events are separated whether they fall inside or outside a polygon around a location, such as a school). In total, these approaches have been consistently used, especially for studies on access to protective features (e.g., treatment facilities) and to exposure to risks (e.g., alcohol outlets). Hot spot or clustering techniques have also been prevalent to varying degrees across all studies. Of particular interest is the emerging use of map analysis in the past ten years. This method is a staple in geographic inquiry and is recognized for its power, not only towards statistical analysis through hypothesis generation, but in its own right as a form of exploratory analysis.

Throughout the extant research, regression approaches made use of spatial data as variables, such as distance or density metrics, but the models themselves were primarily aspatial. The alternative use of spatial models, such as GWR has been relatively late in the time frame of this research. Indeed, one of the first was Yu and colleagues in 2014 who examined the relationship between community pharmacies which sell addictive products, such as tobacco, and the racial composition of the areas in which they are located. Given the proliferation of this technique across the sciences, it is only a matter of time before use of such an explicitly spatial modeling approach is adopted more widely in substance abuse research.

In addition to the topics and methods that are covered in the body of work, in light of the earlier discussion about the importance of scale in GIS, it is also interesting to note the units of data collection and analysis at which this research has been undertaken. Figure 9.6 provides a summary of scale.

Three primary types of scales have been implemented in these studies: census, administrative (other forms of government-defined areas), and functional. Census units are most prevalent and can range from state comparison to as small as blocks. The tract is most frequently

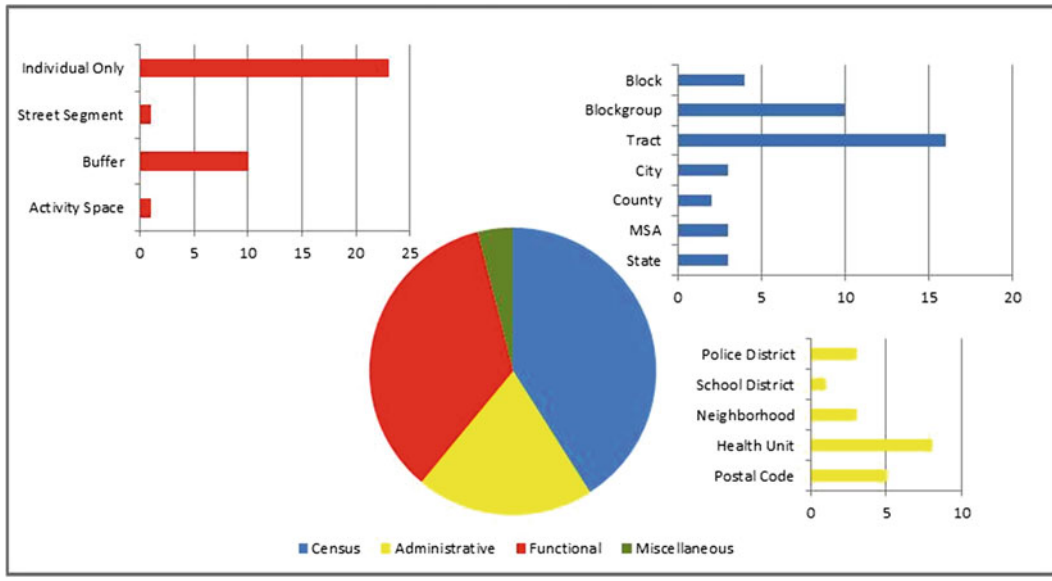

Fig. 9.6 Summary of scales

used due to the amount of SES data available at this level which is often being seen as a proxy for neighborhoods, despite the debate on the appropriateness of this practice. The next most prevalent scale is functional, meaning it is defined based on relevance to some individual or activity. Most of these studies have used only individual level data (geocoded), while others have analyzed features and relationships within set buffer zones around points of interest, within activity spaces of individuals, and even along a street segment. The least implemented scale is administrative, such as postal codes, (but also other non-census zones, such as school districts), other types of districts and health units. Neighborhoods are also included in this category, but only those which are officially defined. This unit of analysis can also be defined by individuals.

Finally, what was also evident from the review of extant literature is the constrained geography of the studies; they are primarily conducted in major metropolitan areas. Of course, this makes sense due to the attention on drugs, alcohol, and violence in cities. However, substance abuse is certainly not limited to these places. Small towns, suburbs, and rural areas are notably understudied with GIS and GT. One explanation is that the skillsets required for geospatial approaches are not as widespread as would be hoped, with more rural areas less likely to have researchers in health units, or proximate centers of higher education that would undertake such studies. As the appreciation for geospatial approaches grows it is possible that more public health researchers and practitioners, especially in these understudied areas will seek out collaborators. From an academic geographer's perspective, these calls for collaboration are often viewed as being exciting as access to health data remains one of the biggest challenges in the field of medical/health geography.

9.4 Emergent Geospatial Approaches to Understanding Substance Abuse

In a review of recent studies, the common trend is *increasing diversity* in the subject areas, types of data and analyses that are utilized. Given these signs of advance in the use of GIS for substance abuse research, it is appropriate to look to "what is next" as this trend expands. Changes are occurring both in the scale of data and content

due to technological advances and understanding the limits of "objective" or "official" GIS. With these changes come the introduction of new forms of analysis and visualization. Overall, there is a need to know more specifically where, when, and why in order to understand and intervene. To this end, there are two related approaches for data collection which are being implemented to provide context and explanation for the existence of hot spots: spatial video and spatial video geonarratives.

9.4.1 Spatial Video (SV)

Spatial video is GPS-enabled high-definition video, which has a coordinate embed per second. This enables the walking or driving path to be captured visually with observable features being geo-located. Typically, this video is displayed in a window that concurrently shows the location of each frame. What is seen can be mapped. This emergent geospatial technology has been employed in a wide variety of settings and for different studies, from patterns of post-disaster damage and recovery, to physical disorder, and environmental health risks.[25] In these applications, the aim has been to link features of the natural or human environment to their real-world location in order to enable mapping and spatial analysis of dynamic or ephemeral data, particularly in challenging environments. The buildings coded in Figs. 9.2 and 9.7 according to their visual assessment are excellent examples of how spatial video can be used to enrich other more traditional data and analyses. The GPS path in Fig. 9.2 shows the spatial video collection route.

9.4.2 Spatial Video Geonarratives (SVGs)

The Spatial Video Geonarrative (SVG) is an environmentally cued narrative where place is used to stimulate discussion about fine-scale geographic characteristics of an area and the context of their occurrence. It is a simple yet powerful approach to enable collection and spatial analysis of expert and resident health-related perceptions and experiences of places. Participants comment about where they live or work while guiding a driver through the area. Four GPS-enabled cameras are attached to the vehicle to capture the places that are observed and discussed by the participant. Audio recording of this narrative is linked to the video via time stamp. A program (G-Code) is used to geotag each word as a point in a Geographic Information System (GIS). Querying and density analysis can be performed on the transcribed text to identify spatial patterns within one narrative or across multiple narratives (Fig. 9.7).[26]

The use of spatial video and SVG are particularly exciting for working with cohorts lacking a permanent address, such as the homeless or sex workers. Both these populations suffer disproportionately from a variety of substance abuse problems, and yet neither fits easily into traditionally geocoded data because there is no known residence. The spatial video can help produce proxies for addresses, while mining the SVG can add context to those maps, including the daily activity patterns, where drugs can be found and used, and how these individuals seek treatment.[27]

9.5 Conclusions

GIS and GT provide new insights into entrenched problems of substance abuse, but can also create new data to identify emerging problems and appropriate interventions. The maps and data generated through these tools have an established history of identifying *where* problems exist, and more recent advances that utilize mixed methods are offering explanations into *why they are there*. It is relatively easy to collect a dataset and then show the hot spots. It is far more difficult to elucidate the mechanisms that formed it.

[25]See, for example, work by Mills et al. (2010), Curtis et al. (2010, 2013), and Curtis and Mills (2011).

[26]See Curtis et al. (2015a, b, 2016).

[27]ibid.

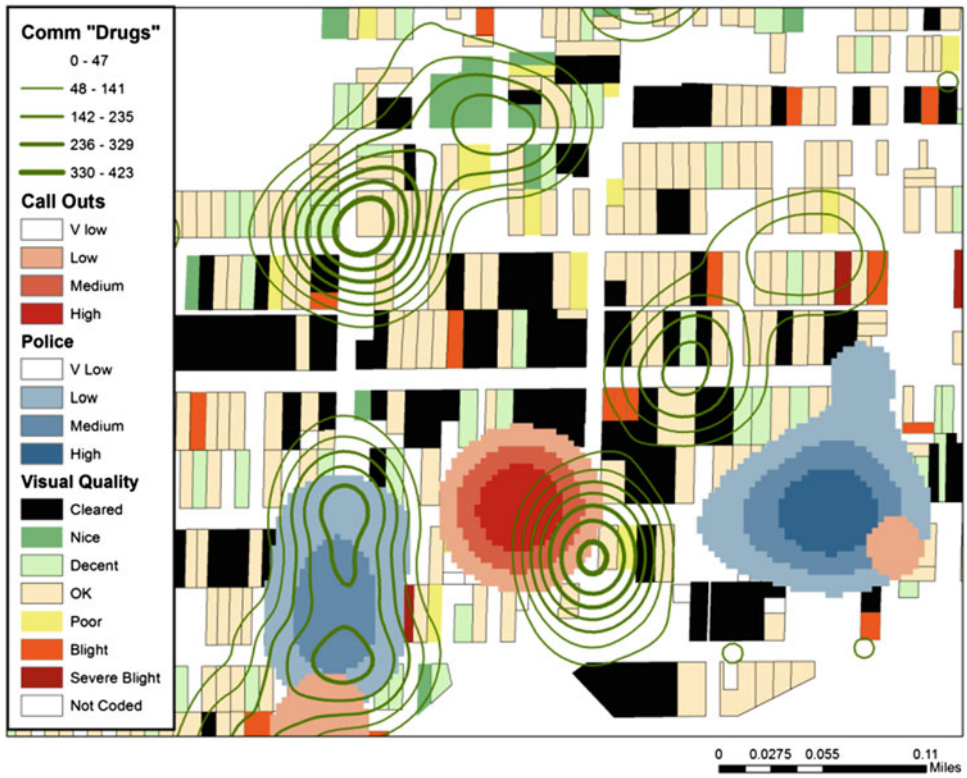

Fig. 9.7 Output from an SVG. *Green contours* hot spot of community member mention of drugs; *Red blobs* hot spots of Police Calls for Service Data related to drugs; *Blue blobs* hot spots of police officer mention of drugs; *Parcel colors* pertain to status of the parcel in a built environment survey

Furthermore, the preponderance of this form of research is cross-sectional, so more questions need to be asked about hot spot spatial-temporal stability or dynamism (e.g., does this hot spot remain unchanged in location, in size, in intensity across all hours, days, weeks, months, years?). Answering such questions is more difficult as it requires more data that may be challenging to adequately collect. However, these answers are instructive to understanding the causal mechanisms and then designing targeted interventions. Furthermore, with such data, interventions can be more rigorously evaluated. In addition to the need for data that are more geographically and temporally granular is the need for more qualitative data from experts and from the people impacted by the problems being studied. Linking explanation to spatial patterns is essential if researchers are going to break apart these hot spots. Emergent advances in geospatial techniques are making this possible, but admittedly more work is required to elucidate methods and to examine the results of their application. This is a time of opportunity for collaboration between geographic information science and public health with the real potential of improving and saving lives.

Acknowledgements National Institute of Justice, Office of Justice Programs. Part of this chapter was supported by Award No. 2013-R2-CX-0004, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice.

References

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis, 27*(2), 93–115.

Baglivio, M. T., Wolff, K. T., Epps, N., & Nelson, R. (2015). Predicting adverse childhood experiences the importance of neighborhood context in youth trauma among delinquent youth. *Crime & Delinquency*. doi:10.1177/0011128715570628

Boulos, M. N., Curtis, A. J., & AbdelMalik, P. (2009). Musings on privacy issues in health research involving disaggregate geographic data about individuals. *International Journal of Health Geographics, 8*(1), 46.

Carlos, H. A., Shi, X., Sargent, J., Tanski, S., & Berke, E. M. (2010). Density estimation and adaptive bandwidths: A primer for public health practitioners. *International Journal of Health Geographics, 9*(1), 39.

Caron, J., Fleury, M., Perreault, M., Crocker, A., Tremblay, J., Tousignant, M., et al. (2012). Prevalence of psychological distress and mental disorders, and use of mental health services in the epidemiological catchment area of Montreal south-west. *BMC Psychiatry, 12,* 183.

Coulton, C. J., Korbin, J., Chan, T., & Su, M. (2001). Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology, 29*(2), 371–383.

Curtis, A., Blackburn, J. K., Widmer, J. M., & Morris, J. G., Jr. (2013). A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Haiti. *International Journal of Health Geographics, 12*(1), 21.

Curtis, A., Curtis, J. W., Porter, L., Jefferis, E., & Shook, E. (2016). Context and spatial nuance inside a neighborhood's drug hotspot: Implications for the crime-health nexus. *Annals of the Association of American Geographers, 106*(4), 819–836.

Curtis, A., Curtis, J. W., Shook, E., Smith, S., Jefferis, E., Porter, L., et al. (2015a). Spatial video geonarratives and health: Case studies in post-disaster recovery, crime, mosquito control and tuberculosis in the homeless. *International Journal of Health Geographics, 14,* 22.

Curtis, A., Duval-Diop, D., & Novak, J. (2010). Identifying spatial patterns of recovery and abandonment in the post-Katrina Holy Cross neighborhood of New Orleans. *Cartography and Geographic Information Science, 37*(1), 45–56.

Curtis, A., & Mills, J. W. (2011). Crime in urban post-disaster environments: A methodological framework from New Orleans. *Urban Geography, 32*(4), 488–510.

Curtis, A. J., Mills, J. W., & Leitner, M. (2006). Spatial confidentiality and GIS: Re-engineering mortality locations from published maps about Hurricane Katrina. *International Journal of Health Geographics, 5* (1), 44.

Curtis, J. W. (2012). Integrating sketch maps with GIS to explore fear of crime in the urban environment: A review of the past and prospects for the future. *Cartography and Geographic Information Science, 39* (4), 175–186.

Curtis, J. W. (2016). Transcribing from the mind to the map: Tracing the evolution of a concept. *Geographical Review, 106*(3), 338–359.

Curtis, J. W., Curtis, A., & Upperman, J. (2015b). Mapping pediatric surge potential after a disaster: Using geospatial techniques for hospital planning and preparedness. In G. J. Kost & C. M. Curtis (Eds.), *Global point-of-care: Strategies for disasters, emergencies, and public health resilience*. Washington, DC: AACC Press.

Curtis, J. W., Curtis, A., & Upperman, J. S. (2012). Using a geographic information system (GIS) to assess pediatric surge potential after an earthquake. *Disaster Medicine and Public Health Preparedness, 6*(02), 163–169.

Curtis, J. W., Shiau, E., Lowery, B., Sloane, D., Hennigan, K., & Curtis, A. (2014). The prospects and problems of integrating sketch maps with geographic information systems to understand environmental perception: A case study of mapping youth fear in Los Angeles gang neighborhoods. *Environment and Planning B: Planning and Design, 41*(2), 251–271.

Dembo, R., Schmeidler, J., Taylor, R., & Burgos, W. (1985). Environmental setting and early drug involvement among inner-city junior high school youths. *Substance Use and Misuse, 20*(8), 1239–1255.

Dent, B., Hodler, T., & Torguson, J. (Eds.). (2009). *Cartography: Thematic map design* (6th ed.). Boston, MA: McGraw-Hill.

Fortney, J. C., Booth, B. M., Blow, F. C., Bunn, J. Y., & Loveland Cook, C. A. (1995). The effects of travel barriers and age on the utilization of alcoholism treatment aftercare. *The American Journal of Drug and Alcohol Abuse, 21*(3), 391–406.

Fotheringham, A. S. (1997). Trends in quantitative methods 1: Stressing the local. *Progress in Human Geography, 21,* 88–96.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. New York, NY: Wiley.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189–206.

Longley, P. A., & Batty, M. (1996). *Spatial analysis: Modelling in a GIS environment*. New York, NY: Wiley.

Maguire, D., Batty, M., & Goodchild, M. (2008). *GIS, spatial analysis, and modeling*. Redlands, CA: ESRI Press.

Mason, M., Cheung, I., & Walker, L. (2004b). Substance use, social networks, and the geography of urban adolescents. *Substance Use and Misuse, 39*(10–12), 1751–1777.

Mason, M., Cheung, I., & Walker, L. (2009). Creating a geospatial database of risks and resources to explore urban adolescent substance use. *Journal of Prevention & Intervention in the Community, 37*(1), 21–34.

Mason, M. J., Cheung, I., & Walker, L. (2004a). The social ecology of urban adolescent substance use: A case study utilizing geographic information systems. *Journal of Primary Prevention, 25*(2), 271–282.

Mazumdar, S., Rushton, G., Smith, B. J., Zimmerman, D. L., & Donham, K. J. (2008). Geocoding accuracy and the recovery of relationships between environmental exposures and health. *International Journal of Health Geographics, 7,* 13.

Messiah, A., Navaline, H., Davis-Vogel, A., Tobin-Fiore, D., & Metzger, D. (2003). Sociodemographic and behavioral characteristics associated with timeliness and retention in a 6-month follow-up study of high-risk injection drug users. *American Journal of Epidemiology, 157*(10), 930–939.

Milam, A. J., Furr-Holden, C. D., Bradshaw, C. P., Webster, D. W., Cooley-Strickland, M. C., & Leaf, P. J. (2013). Alcohol environment, perceived safety, and exposure to alcohol, tobacco, and other drugs in early adolescence. *Journal of Community Psychology, 41*(7), 867–883.

Mills, J. W. (2009). Spatial decision support in a post-disaster environment: A community-focused approach. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44* (1), 17–31.

Mills, J. W., Curtis, A., Kennedy, B., Kennedy, S. W., & Edwards, J. D. (2010). Geospatial video for field data collection. *Applied Geography, 30*(4), 533–547.

Monmonier, M. S. (1996). *How to lie with maps*. Chicago, IL: University of Chicago Press.

Muth, S. Q., Potterat, J. J., & Rothenberg, R. B. (2000). Birds of a feather: Using a rotational box plot to assess ascertainment bias. *International Journal of Epidemiology, 29*(5), 899–904.

O'Loghlen, S., Pickett, W., & Janssen, I. (2011). Active transportation environments surrounding Canadian schools. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique, 102*(5), 364–368.

Openshaw, S. (1996). Developing GIS-relevant zone-based spatial analysis methods. *Spatial Analysis: Modelling in a GIS Environment*, 55–73.

Pang, T. T. P., & Lee, S. S. (2008). Measuring the geographic coverage of methadone maintenance programme in Hong Kong by using geographic information system (GIS). *International Journal of Health Geographics, 7,* 5.

Rockwell, R., Jarlais, D. D., Friedman, S. R., Perlis, T. E., & Paone, D. (1999). Geographic proximity, policy and utilization of syringe exchange programmes. *Aids Care, 11*(4), 437–442.

Tanjasiri, S. P., Lew, R., Mouttapa, M., Lipton, R., Lew, L., Has, S., et al. (2013). Environmental influences on tobacco use among Asian American and Pacific Islander youth. *Health Promotion Practice, 14*(5 Suppl), 40S–47S.

Walker, L. R., Mason, M., & Cheung, I. (2006). Adolescent substance use and abuse prevention and treatment: Primary care strategies involving social networks and the geography of risk and protection. *Journal of Clinical Psychology in Medical Settings, 13* (2), 131–139.

Wiehe, S. E., Kwan, M., Wilson, J., & Fortenberry, J. D. (2013). Adolescent health-risk behavior and community disorder. *PLoS ONE, 8*(11), e77667.

Yu, D., Morton, C. M., & Peterson, N. A. (2014). Community pharmacies and addictive products: Sociodemographic predictors of accessibility from a mixed GWR perspective. *GIScience & Remote Sensing, 51*(1), 99–113.

All Mixed Up: Considerations in Blending Qualitative and Quantitative Components in Substance Abuse Research

10

Sheryl L. Chatfield and Jeffrey S. Hallam

10.1 What Do We Mean by Mixed Methods?

It is likely that readers of this chapter have a preconceived notion of what the term *mixed methods* means; it also possible that there is variation among those meanings. In an article helpfully entitled "Toward a Definition of Mixed Methods Research," Johnson et al. (2007) compiled a definition based on the input of "leading mixed methods research methodologists" (p. 118). The following description has been adopted, shown below, as a guide for the material in this chapter.

> Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches … for the broad purposes of breadth and depth of understanding and corroboration (Johnson et al. 2007) (p. 123).

10.1.1 Other Terms

We adopt Patton's (2002) definition of *qualitative* research as that concerned with data derived from interviews, observation, or document analysis. We consider *quantitative* research as that focused on numbers, whether the numbers are presumed to be true representations (e.g., chronological age or count of time something occurred) or used as proxies (e.g., use of 1 through 5 to stand for responses choices on a survey). Quantitative research encompasses not just descriptive and inferential statistics, but also other arithmetic results.

When merging multiple types or bouts of research, some authors use alternative terms such as *multiple methods*, *multimethods, or combined methods.* These terms may be used as alternatives to mixed methods (e.g., Teddlie and Tashakkori 2009; Gorard and Taylor 2004), or to describe specific ways of mixing methods or studies (e.g., Morse and Niehaus 2009; Hesse-Biber 2010). To prevent ambiguity or confusion, we refer to the research designs of interest in this chapter only as mixed methods.

10.2 Paradigms

Teddlie and Tashakkori (2009) defined a *paradigm* as broad worldview that underlies any given research study. Mixed methods researchers frequently contrast what Teddlie and Tashakkori (2009) called the "constructivist" and "postpositivist/positivist" paradigms. The former

S.L. Chatfield (✉) · J.S. Hallam
Department of Social and Behavioral Sciences, College of Public Health, Kent State University, Kent, OH, USA
e-mail: schatfi1@kent.edu

J.S. Hallam
e-mail: jhallam1@kent.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_10

approach is associated with the identification and use of commonalities or themes applicable to one or a few to facilitate deeper understanding of individuals' experiences. Qualitative research is often fit within this paradigm. The post-positivist/positivist is associated with use of scientific methods to describe or confirm findings that can be generalized to larger groups; this is generally associated with quantitative methods.

Researchers' opinions regarding the impact of paradigm on method mixing can generally be categorized in one of the following four ways: (1) the paradigms are incompatible such that qualitative and quantitative methods cannot be mixed, (2) the paradigms are incompatible, so mixed methods research needs to be guided by an alternative paradigm, (3) the paradigms are not necessarily compatible, but with care both may be accommodated within a single study, or (4) a decision to collect a certain type of data does not require that a researcher align him or herself with a specific paradigm.

We believe that ample high quality mixed methods research studies have been conducted and published to overrule the incompatible paradigms argument. While it is not our intent to argue for any one of the three other viewpoints, we believe that researchers should be prepared to describe and defend their design decisions and reviewers should request clarifying information when not provided. Following, we provide descriptions and sources for designs based on categories 2, 3, and 4.

10.2.1 Pragmatism: The Other Paradigm

Pragmatic research can be briefly described as results focused. American educator and writer John Dewey (1859–1952) is viewed as an influential advocate of the pragmatic approach (Teddlie and Tashakkori 2009). Greene and Hall (2010) described the ideal outcome of a pragmatic approach as providing "contributions to workable solutions to important problems." A pragmatic approach encourages researchers to use the methods from both the qualitative and quantitative tool belts that offer the best chance of responding to the question or problem of interest. Additionally, pragmatists remain open to the possibility that the appropriate tools may change at various phases of a research project, although the focus remains generally the same. Teddlie and Tashakkori (2009) recommended the pragmatic approach as the approach best suited to mixed methods research.

10.2.2 The Dialectic Alternative: Paradigm Accommodation

The "dialectic" alternative (Greene and Hall 2010) provides a means for researchers to accommodate the incompatible nature of the paradigms. According to Greene and Hall, this approach "actively welcomes more than one paradigmatic tradition" and allows researchers to blend qualitative and quantitative methodologies without divorcing the methods from associated worldviews. Greene and Hall noted that the outcome of dialectic research might provide "meaningful engagement with differences that matter." With regard to mixed methods approaches to evaluation, Patton (1997) noted: "a flexible, responsive evaluator can shift … between paradigms … within a single setting … and can help adherents of either paradigm interpret data in more than one way" (p. 296). A researcher using the dialectic stance may generally be likely to accept, engage with, explore, and present differences in findings resulting from different methods. The pragmatist, on the other hand, might be more likely to try to find common ground, even if it requires reshaping or deemphasizing findings that are inconsistent.

10.2.3 Design-Focused: A Natural Approach

Greene and Hall (2010) used the term "aparadigmatic" to describe research "more directly informed by theory, context, and/or ideology." We, however, prefer to think of this

way of proceeding as *design-focused,* and look to Gorard (2010) and Gorard and Taylor (2004) for further discussion. According to Gorard and Taylor, when "combining methods," researchers should focus on research questions rather than personal preferences, work to acquire the skills necessary be able to use multiple methodologies and methods, and be prepared to address research questions "by whatever means it takes." Other key ideas Gorard (2010) described include: (1) acceptance that mixed methods is a natural approach and is reflected in many real-world decisions, (2) understanding that there are similarities as well as differences between qualitative and quantitative methods, and (3) accepting the ethical obligation, in particular for funded researchers, to focus on "quality of research… robustness of findings and the security of the conclusions drawn" (p. 247) rather than displaying favoritism or preference for one methodology or paradigm over another.

10.3 Why Use Mixed Methods?

According to Miles et al. (2013), it is necessary "to face the fact that numbers and words are both needed if we are to understand the world" (p. 42). We believe that mixed methods research designs offer several strengths that are of particular importance to substance abuse research. These include:

- **Mixed methods research designs can help overcome some of the challenges in conducting substance abuse research**. Ethical concerns impact inclusion of control and comparison groups, while the nature of substance abuse research may, in general, encourage purposive rather than random sampling. Either of these issues can have a negative impact on a quantitative/statistical design, while the collection of complementary data through qualitative methods may help to validate, or at least add more dimensions to the findings.
- **Mixed methods research designs can more comprehensively address research questions as opposed to either qualitative or quantitative approaches implemented alone**. For example, a finding of statistical significance and meaningful effect size from quantitative research might suggest that a given treatment is associated with a difference in response; incorporation of observational or interview research conducted concurrently might help to explain why a treatment is effective or clarify some of the differences among responses.
- **Mixed methods research designs are applicable to various settings or contexts**. Substance abuse research may focus on contexts ranging from one-on-one therapy sessions to program evaluation. Through the use of randomization techniques (discussed in Sect. 10.4), statistical analyses can be conducted for single-case or small samples. Furthermore, the need for comprehensive information argues for the use of mixed methods in most, if not all, program evaluation research.
- **Mixed methods research designs are better able to handle the complexities within substance abuse research**. Substance abuse is frequently coexisting, or related to other subject matter of social and behavioral research, including mental illness, crime or deviant behavior patterns, risky sexual behaviors, homelessness, and/or abusive relationships. These multiple areas of concern benefit from a multitiered approach to research.
- **Mixed methods research helps to bridge the gap between academics and practitioners**. Skillsets and access to resources may result in differences in ability or focus between academics and practitioners. Therapists' access to and focus on each client's treatment trajectory provides an ideal focus for qualitative research. Academic researchers, especially US-educated researchers, may have greater comfort level and experience

analyzing quantitative results, such as those obtained from assessing outcomes from multiple clients serviced by a clinic, program, or health system. The combination of detailed, qualitatively analyzed findings from a small number, with statistical results from a large group, is an essential mixed methods research design.

- **Mixed methods research is broad and flexible, and is so suited for a dynamic topic like substance abuse**. Legal and societal responses to substance abuse have changed over time, and it is reasonable to anticipate that changes will continue. Development of expertise in mixed methods designs makes researchers more able to respond to changing priorities.
- **Mixed methods research is fundable**. A report provided by the National Institutes of Health (NIH) online RePORTER suggested that funding for current mixed methods research projects by the National Institute for Drug Abuse (NIDA) totals more than $15 million, with an average project allocation size over $300,000 (NIH 2016).

10.4 Categorizing Designs

Based on our definition, any mixed methods research project includes at least one qualitative and one quantitative component. Researchers have described a multitude of combinations using these elements; for example, Creswell and Plano Clark (2011) provided a table that summarizes 15 typology schemes, each of which includes multiple variations. In this section, we provide four alternative views of mixed methods design, presented in chronological order of publication, to represent a range of variations that are particularly applicable to substance abuse research topics.

Greene et al. (1989) provided one of the earlier means of categorizing mixed methods designs. According to Greene et al. mixed methods research designs can be distinguished based on the order of the methodologies used (sequential or concurrent), the emphasis placed on each (placing higher priority on one portion or considering each as equal), whether there is one or more than one specific focus, and whether the underlying paradigm (see 1.2) is the same or different for each methodology. Greene et al. further distinguished research designs based on the end goal of the mixed methods approach (e.g., confirming findings versus augmenting or expanding findings).

Hesse-Biber (2010) described her approach to mixed methods research as a qualitative approach, characterized by empathy for participants and respect for "human subjectivity." According to Hesse-Biber, distinctions among designs within qualitatively focused mixed methods are based on whether research is sequential or concurrent, and at what point integration of these different methods occurs.

Morse and Niehaus (2009) distinguished designs based on timing (sequential or concurrent) and emphasis. According to Morse and Niehaus, mixed methods research, by definition, must contain a primary and secondary component. Findings from these components comprise a single study, which prohibits researchers from writing or presenting the findings separately.

Miles et al. (2013) offered three specific ways of combining methods. The first is in converting qualitative data into numerical or ordered expressions. The second is "linkage," or the comparison of two types of data from the same source. The third combination occurs at the design level. These authors provide some examples (e.g., continuous qualitative data collection punctuated by bouts of survey research), while noting that many "more complex" alternatives exist.

10.4.1 "Designer" Approaches

Essentially any qualitative methodology can form the qualitative component of a mixed methods research study. It is our observation that the qualitative component of much mixed methods research reflects what is often referred to as a generic or descriptive approach. We expect that this will continue to be one of, if not the most common qualitative methodology in

mixed methods research, especially for researchers who have not had a great deal of training or exposure to a wide range of qualitative methodologies. Discussions on descriptive, or otherwise known as generic approaches to qualitative inquiry, can be found in Caelli et al. (2003) or Sandelowki (2000).

Some traditionally qualitative designs can accommodate quantitative, as well as qualitative data, to reflect potential candidates for expansion into mixed methods research. These designs include grounded theory, case study, and ethnography.

Glaser and Strauss (1967) defined grounded theory as the "discovery of theory from data systematically obtained from social research" (p. 2). Glaser and Strauss also saw "no fundamental clash between the purposes and methods of qualitative and quantitative methods or data … [i]n many instances, both forms are necessary" (p. 17–18). By its nature, grounded theory presumes an evolving approach and the value of an alternative methodology might become apparent at any point after initial data are analyzed. However, we want to point out that midstream modification of a research design may create issues with institutional review boards or other oversight organizations, in addition to complicating evaluation plans.

Researchers use case study research designs to gain better understanding of a "contemporary phenomena in depth and within its real-life context" (Yin 2009, p. 18). Multiple or collective case study designs may be used to assess more than one item, whether the item of interest is individuals, groups, or multiple sites within a program. By its nature, the case study design presumes collection of multiple data types; according to Yin (2009), case study research regularly "goes beyond being a type of qualitative research, by using a mix of qualitative and quantitative evidence" (p. 19).

Ethnography refers to a type of cultural exploration that has its roots in anthropology and sociology. While some well-known twentieth century urban ethnographies (e.g., Liebow 1967; Stack 1975) are presented for the most part as descriptive reflections, there is generally an element of quantitative data present in the form of demographic or trend data. Spradley's (1970) exploration of the lives of men who were repeat violators of public drunkenness statutes incorporates several quantitative elements, such as a report of the results of a survey completed by 101 volunteer participants.

10.4.2 More Than Two

Several of the aforementioned design descriptions suggest a two-part (i.e., one qualitative and one quantitative component) research study. However, as Miles et al. (2013) suggested, there are myriad other design alternatives. Additionally, as suggested in the grounded theory discussion, the need for another type of data may become apparent during the course of a research study; Creswell and Plano Clark (2011) referred to this as an "emergent design," in contrast to a "fixed design," in which all elements are planned prior to implementation. Therefore, researchers should not limit their thinking to a single qualitative method and a single quantitative method. For example, observational research might be conducted to help design a survey instrument. Following administration of the instrument, the researcher might wish to debrief a select group of survey respondents. Based on the results of the debriefing, the researcher may wish to refine and re-administer the survey. Time, resources, and other considerations may prevent spontaneous implementation of such a design, but this example is included so readers may begin to visualize increasingly complex designs.

10.4.3 Common Elements

It is not the intent of this chapter to direct readers to a specific research design, but there are some commonalities among the alternatives provided that are worth highlighting. First, researchers should decide whether a wholly emergent or grounded theory approach is warranted and possible. If so, it may only be necessary to

specify the research question and initial data collection strategies. We do not expect that this is often going to be the case, and therefore recommend interested readers to Glaser and Strauss (1967) or Glaser (1978) for more details regarding this approach. In most other instances, it is necessary to determine whether data collection from qualitative and quantitative methods will be sequential or concurrent. It may be helpful to determine if and how priority will be given to one method or data type over another or to remain flexible about this part of the design, though the authors believe it is possible to assign equal priority.

10.4.4 Planning and Implementing Mixed Methods Research

There is no practical way to anticipate or address all potential substance abuse research questions that might be addressed through mixed methods designs. Instead, Table 10.1 includes increasingly complex example scenarios, possible paradigm associations, categorization from the design elements discussed above, and potential methods.

10.4.5 Analysis and Presentation of Findings

In an assessment of more than 200 social science research reports described by the authors as mixed methods, Bryman (2006) determined that the most common was a combination of structured (survey) interviews and semi-structured (qualitative) interviews. Bryman (2007) also interviewed researchers about the integration of mixed qualitative and quantitative research findings. Commonly identified challenges included a lack of good examples, researchers' tendency to emphasize the type of research he/she was most comfortable with, and a perceived incompatibility between data types. Bryman (2007) observed that parallel, rather than truly integrated presentations, were common.

Bryman's (2007) results suggest that the number of researchers who are confident in their ability to analyze findings from both qualitative and quantitative research is very likely smaller than the number of those who are confident of their skills in one or the other. For this reason, much mixed methods research benefits from a partnership or group approach to manage skill differences and assist in quality control. As with other elements of mixed methods research design, several authors (e.g., Creswell and Plano Clark 2011; Morse and Niehaus 2009) have provided guidance, typologies, and even rules for data integration during analysis and integration of findings for presentation. The following guidelines are neither exhaustive nor inflexible, but provide a starting point for novice mixed-methodologists, as well as considerations for more experienced researchers.

10.4.5.1 First Steps in Analysis

We believe that it is necessary to begin analysis of mixed methods findings by beginning with standard approaches to the individual components that comprise the study. It is beyond the intent of this chapter to provide detailed directions for either qualitative or quantitative data analysis.[1] Several of the more common qualitative analysis software programs also provide some manuals or online guidance, although it is necessary to have some understanding of initial coding to reap adequate benefits from the software. Saldaña (2013) detailed the use of Microsoft Word® as an alternative to some of the dedicated analysis software programs. It is more difficult to make recommendations for quantitative analysis resources because of the range of potential models and the more profound dependence on software. Many multivariate methods texts (e.g., Hair et al. 2009; Tabachnick and Fidell 2012) contain directions for the more commonly used statistical software packages, such as SPSS or SAS. There are also several

[1]For a clear and comprehensive explanation of qualitative data analysis that includes both generic and "designer" approaches, suitable for both beginners experienced analysts, see Saldaña (2013). A somewhat more systematized approach is detailed in Miles et al. (2013).

Table 10.1 Mixed methods substance abuse research design examples

Research focus	Paradigm	Timing	Priority on results	Methods	Integration	Goal
Client assessment	Dialectic	Concurrent	Quantitative	Results of standardized instruments; notes from observations	During analysis	Develop an individualized treatment plan
Client evaluation	Dialectic	Sequential	Qualitative	Semi-structured interview followed by randomized observations (see Sect. 10.4)	During analysis	Explore client's perception of treatment process; use observations to verify (or question) interview findings
Treatment evaluation	Pragmatic	Concurrent	Quantitative	Meta-analysis; Meta-synthesis (synthesis of qualitative research reports)	After both analyses are complete	Report a comprehensive review of research on the treatment of interest
Program evaluation	Pragmatic	Concurrent	All	Key informant interviews; surveys of outcomes; surveys of satisfaction; observation of operations/staff	Ongoing throughout the process	Produce a comprehensive picture of operations and outcomes
Development of intervention	Design-focused	Sequential	All	Meta-analysis followed by group interview followed by factorial survey (see Sect. 10.4) followed by pilot test of intervention	Progressive (each component is informed by all prior data)	Implement an effective intervention that is attractive to clients

texts written for specific packages, such as R (e.g. Dalgaard 2008).

Because the ultimate goal of analysis is to present some connection between qualitative and quantitative findings, whether it is via a parallel presentation, a sequential presentation, or a more integrated presentation, neither qualitative nor quantitative analysts can complete their parts of the process entirely in isolation. It is helpful to keep in mind the rationale for mixing methods in the first place, which is usually some version of verification or augmentation. We advise that mixed methods researchers adopt the practice, more generally associated with qualitative inquiry, of writing throughout the research project. There is value in writing from the planning stages of a project onward, leaving the necessary gaps to insert findings or analysis. If researchers are disciplined in this practice, changes or alterations can be incorporated into the written report as they occur, which may prevent later difficulties resulting from less than perfect memory.

10.4.5.2 Integration and Presentation Alternatives

Once method-specific analyses are complete, we recommend that researchers compare the findings and consider both textual and graphical representations. Creswell and Plano Clark (2011) provided a model research paper with a separate chapter devoted to each qualitative and quantitative research finding. The authors agree with Bryman (2012) by, in general, *not* recommending this approach. However, if two or more components of a single study are directed at different phenomena, a direct comparison between or among all may not be possible. In this circumstance, findings from each component should be presented sequentially, in priority order when applicable, and strive to build some transitional bridge between the components in the text of the report. A cohesive graphical presentation may not be possible if the results are too disparate.

There are multiple other patterns that facilitate comparison may emerge following analysis. Two of these are findings that are either conflicting or identical. For either of these alternatives, it is recommended that researchers alternate descriptions of the primary findings and the conflicting or supporting findings, illustrated by some type of parallel presentation, such as two bulleted columns. The following flowchart presents some guidelines for both written and graphical presentation of data that are not so clearly in conflict or agreement, and instead reflect a less definitive degree of congruence or contrast. For visual presentations, we refer to some of the standard graphics available within the Microsoft Office® SmartArt, though researchers should invest the time necessary to explore ways to modify these or create appealing and situation appropriate visual representations of data. Note that in each instance it is up to the researcher, based on research questions, design elements, or character of the findings, to determine which component is considered primary for the purposes of presentation of data (Fig. 10.1).

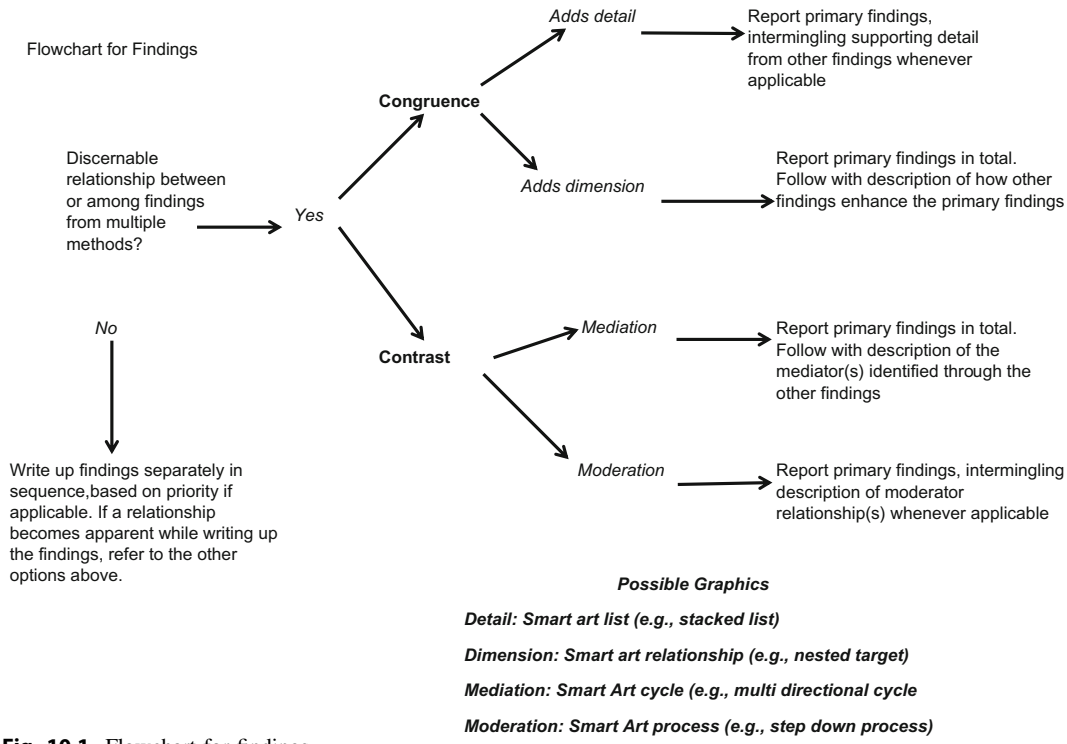

Fig. 10.1 Flowchart for findings

10.4.5.3 Paradigms Revisited: The Discussion Section

In the prior section, we provided recommendations for writing up and presentation of the essential findings. Although for much quantitative analysis, tables or written descriptions of results may be largely self-explanatory, qualitative researchers, on the other hand, may strive for levels of description that approach or attain "interpretive explanation … [through] transformation of data" (Sandelowski and Barroso 2003, p. 914). The discussion section in a mixed methods research report may lean toward one or the other of these extremes, may attempt to reflect both, or may present a view of compromise. While in many instances some separation in the presentation of findings was advised, it is preferred that the discussion section of a research report, or the discussion segment of a presentation, reflect a greater degree of integration and cohesion, even if the findings are in conflict. One exception to this is mixed methods research that consists of components directed at different phenomena. In this instance, researchers are encouraged to consider what unifying factors exist among the components; if there are no ways to tie the findings together, then the project may reflect not one, but two or more essentially independent research studies. Ultimately, emphasis within the discussion section may vary depending on the researcher's orientation or paradigm view.

Researchers who have adopted a pragmatic focus will emphasize the elements within the findings that most directly address the research question or concern. This means that findings that are less helpful may not be reported. On the other hand, when analysis of one component reveals conflicting findings, this may point to the need for further research and/or the use of different or additional methods. Conflicting findings should always be reported.

Researchers who have adopted a dialectic approach may wish to report quantitative findings definitively and with a focus toward responding to the research question (e.g., descriptive statistics, effect sizes) and qualitative findings more interpretively, perhaps with incorporation of further exploration of literature conducted after analysis. It is the authors' belief that researchers should find some way to 'marry' these results in the discussion question, if necessary by revisiting the level of themes or codes in the qualitative analysis for comparison with quantitative results.

Researchers who have adopted a designed-focused approach, as described above, will want to borrow the research question focus from the pragmatic approach, while also considering thorough assessment of both qualitative and quantitative data as unique components, and as elements of a single study. The discussion section in a design-focused study should address the research questions, although it may be more likely than other approaches to include further questions or ambiguities in the findings.

In general, the discussion section offers the researcher the opportunity reconsider the research question or concern, to assess the worth of the component findings in addressing this, to revisit the research cited in the background or review of the literature, and to consider contributions of the research to existing or new theoretical directions. If a research study contains more than two components, regardless of the timing (concurrent, sequential, or some combination of the two), all elements should be revisited in the discussion section.

In closing, it is important to point out that data reduction is a shared goal of both qualitative and quantitative approaches. Given that a mixed methods research study will most likely yield more findings than are practical or desirable to report, researchers should remain aware that by trying to fit too much into a paper or presentation, they risk losing sight of the purpose of the research, as well as losing the attention of the audience who could benefit. As such, researchers should not necessarily discard data that is of interest but that does not address the current research question or concern, but instead consider how these additional findings might provide incentive or justification for other research projects.

10.5 Examples of Mixed Methods Research

This section provides a small number of examples of both funded and published research, and discusses each in relation to the foundational information provided in the first two sections of this chapter. Selections were made to represent a variety in both substance abuse research approach and focus.

10.5.1 Funded Research

According to the NIH RePORTER, as of the spring of 2016, a total of 47 active substance abuse research projects were funded by NIDA and described as mixed methods (NIH 2016). These projects represent a range of locations, populations of interest, specific subject matter, and research methods. In Table 10.2, six current and recent examples are included that demonstrate the range of funded projects through this single agency. The table details related to elements of the foundational knowledge provided in the first two sections of this chapter when this information was readily available in the RePORTER record. More explicit detail about these or other mixed methods research projects, funded through any NIH agency (e.g., National Institutes of Mental Health, National Institute on Aging), and including citation information for resultant publications, is available via the NIH RePORTER.

Additionally, the NIH office of Behavioral and Social Science Research has commissioned guidelines (Creswell et al. 2011) for researchers in those areas who want to pursue mixed methods approaches. These guidelines are available for download at: http://obssr.od.nih.gov/scientific_areas/methodology/mixed_methods_research/index.aspx.

10.5.2 Published Research

We conducted general database searches for published mixed methods research reports devoted to substance abuse topics. As might be assumed based on the publication dates of several of our technical references, the number of published research reports has dramatically increased in recent years. Using a publication date of 1998, our search for "mixed methods" and "substance abuse" produced 854 peer-reviewed publications; after changing the date limiter to 2004, the number of reports decreased only slightly to 842, and a search from 2010 to 2016 brought up 699 hits. Summarized below are four research reports that demonstrate variety in subject matter and methods.

- **Longitudinal assessment of impact of substance abuse treatment intervention**
  Rayburn (2013) explored homeless individuals who had sought treatment for substance abuse. The author used treatment center records to compile demographic, diagnostic, and mortality information, performed an in-depth public records search of a sample of treated clients to assess posttreatment criminal history, and conducted in-depth interviews with alumni of the treatment center. Although findings from qualitative and quantitative components were published separately, we highlight the cited article because it provides detailed descriptions of both the qualitative and quantitative methodologies.
- **Exploration of factors contributing to high dropout rates in residential drug abuse treatment from both client and staff standpoint**
  Brener et al. (2010) combined demographic and attitudinal data collected via survey, in addition to a thematic analysis of interview transcripts, to assess the impact of client perceptions of staff attitudes on heroin users' completion of residential drug rehabilitation treatment programs. Findings from both survey and interview components suggested that clients perceived discriminatory attitudes from staff, and that these attitudes were a deterrent toward program completion. Interviews with staff elicited some recommendations to improve client perceptions.

Table 10.2 Current or recently NIDA funded research

Research focus	Researcher organization	Timing	Methods	Goals	Other comments
Prescription opioid abuse	University of Colorado	Multiple sequential components	Interviews and focus groups; develop intervention; conduct randomized controlled trial of intervention	Develop and pilot test intervention	Qualitative data collection used to inform later stages
Consequences of heroin use based on availability and purity	University of California, San Francisco	Concurrent	Secondary analysis of government databases; interviews and ethnography	Compare reported health outcome data with observations and interview findings from current heroin users	Two types of data used in combination to develop comprehensive understanding
Drug use in clubs by Asian American men who have sex with men	Scientific Analysis corporation	Concurrent	Interviews, ethnography; survey instrument for demographic information	Assessing drug use and other risky behavior patterns	Continuation of prior research; stated focus on qualitative methods
Substance abuse screening process	New York University School of Medicine	Multiple sequential components	Psychometric testing of instrument; interviews	Modify existing instrument; assess psychometric properties and physician receptiveness to screening process; ultimate goal is to increase use of screening during physician visits	Interviews conducted with providers and patients
Examination of the impact of healthcare reform on drug addiction and HIV services	University of California, San Francisco	Multiple concurrent and sequential components	Pre and post test cohort comparison; followed by longitudinal component with one group; secondary data analysis, interviews with key informants	Assess changes in services and demands after implementation of the Affordable Care Act	Similar in structure to evaluation research
Increasing availability of substance abuse treatment via primary care	Rand Corporation	Multiple concurrent components	5 year randomized controlled trial; interviews, observations, document review; secondary analysis of patient and provider records	Compare outcomes from evidence-based substance abuse treatment protocols suitable for delivery in primary care settings	Multisite, multi-methods research study with multiple qualitative and quantitative components

- **Secondary data analysis to assess factors in substance abuse relapse and recovery**
  VanDeMark (2007) conducted secondary data analysis on a randomly selected sample from a previously conducted national survey of women with substance abuse histories to assess the role of social and instrumental support in relapse and recovery. Qualitative findings were reported via themes developed, in part, on frequency of occurrence; these were compared to the results of logistic regression analysis designed to assess the relative role of various types of support in relapse. The author concluded that both qualitative and quantitative findings argued for the importance of social connections in lasting recovery.
- **Three case studies focusing on reduction of HIV risk among injection drug users**
  Wagner et al. (2012) represented and reevaluated information from three previously published mixed methods case studies. The authors demonstrated how in each instance, the qualitative and quantitative components yielded "discordant findings," yet through the integration process they achieved "more nuanced understanding of the issues than may have been achieved through the use of a single method." Wagner et al. cautioned against reliance on any single paradigm and advised that researchers consider analysis strategies in advance to manage conflicting results from different components in mixed methods research studies. This article is highlighted in particular as an exemplar of a published mixed methods report that provides comprehensive and balanced information about the approach in general and a thoughtful, reflective discussion of research findings.

10.5.3 Comments

Our review of published research provided support for Bryman's (2006) findings that published, mixed methods research articles are likely to describe the use of qualitative individual or group interviews, combined with a quantitative component represented by survey research. The qualitative component was not often described as belonging to a specific qualitative methodology, although elements of qualitatively driven grounded theory analysis (e.g., open and axial coding) were frequently described. This may be due to the fact that grounded theory texts (Strauss and Corbin 2008) include more vivid descriptions of the coding process than some general and some method-specific qualitative texts. Those readers who are not using a grounded theory methodology are encouraged to consider coding alternatives such as descriptive, emotion, in vivo (use of source wording to create codes), or other schemes described by Saldaña (2013).

In both our examples and other reviewed research articles, we noted that findings from the multiple (usually two) components of mixed methods studies were generally reported sequentially, and integration consisted of one or more summarizing statements in the discussion. Additionally, in several instances, findings were reported in separate published articles. Pressure from both funding agencies and tenure granting institutions likely encourage this practice, although this segregation of publication or reporting may also result from the researchers' lack of confidence for integrated analysis.

It is encouraging to see several studies that appeared to reflect an integrated approach, beginning with project initiation, in our review of currently funded research, much of which consisted of more than a single qualitative and quantitative component. This may, in part, be attributable to complex subject matter; several funded studies proposed assessment of multiple behaviors (e.g., drug use and HIV risk) or multiple populations (e.g., adolescents and their parents). It is hopeful that the increasing sophistication in funded mixed methods substance abuse research studies reflects increasing researcher competence, as well as increasing enthusiasm for the approach.

10.6 Considered Combinations

In the final section of this chapter are potentially effective approaches to mixed methods research that take advantage of some of the strengths of both qualitative and quantitative research. To begin, a factorial survey is a quantitatively driven process that can be easily expanded to incorporate a meaningful qualitative element. Discussed thereafter is a particularly effective approach for client-focused substance abuse research: a combination of a phenomenological approach to interview research with supporting quantitative data collected via single-case or small *n* randomization. For the final alternative, ways to incorporate nontraditional data types, including technology-driven data, into substance abuse research are considered.

10.6.1 The Factorial Survey

The factorial survey is a quantitative research approach described by Rossi and Anderson (1982), which was developed for social and behavioral research. A factorial survey instrument consists of a series of similar short scenarios, also known as vignettes, which vary slightly in detail. Participants rate or score multiple variations of the same scenario and the results are used in a regression analysis to assess the weights of factors in decision-making, attitudes, or intentions. Wallander and Blomqvist (2009) used factorial surveys to assess whether Swedish social workers would recommend inpatient or outpatient substance abuse treatment under varying conditions. In addition to clinical decision-making, we believe that factorial surveys are well-suited for intervention planning and assessment of attitudes or judgments.

A typical factorial survey instrument includes vignettes composed of several characteristics, which represent the independent variables, and some method of indicating an outcome variable, whether by score (e.g., *using a scale of* 0–10*, rate the attractiveness of this treatment alternative),* ordered response (e.g., *strongly agree, agree, neutral,* etc.), or probability (e.g., *on a scale of* 0–100*, how likely is someone to inject heroin given the circumstances described above*). When each respondent is given multiple scenarios to rate or rank, a large number of observations can be compiled from a smaller number of respondents. The characteristics or dimensions are generally derived from existing literature, although items of researcher interest may be included; this type of survey offers a low-cost, low-risk way of exploring many variables. For inpatient substance abuse treatment planning, vignettes might address variables, including duration of program, frequency, duration and structure of therapy sessions, provision of recreational activities, staff qualifications, characteristics of other clients, and location of facility.

As explicit detail regarding the creating and randomizing of instrument elements is beyond the scope of this chapter, readers are referred to Rossi and Anderson (1982), Jasso (2006), and Wallender (2009) for detailed descriptions of the method, Dulmer (2007) and Hox et al. (1991) for discussion of statistical analysis, and Auspurg et al. (2009) and Sauer et al. (2010) for discussions of construction of vignettes.

Ganong and Coleman (2005, 2006) extended the factorial survey instrument through inclusion of one or more qualitative or open response items. An extension of the instrument is recommended to incorporate a substantial qualitative component through the use of multiple open response items, allowing respondents an opportunity to disclose or discuss their thought process when scoring or rating vignettes. Potentially interesting and enlightening data are obtained when participants are provided with space to provide their own versions of vignettes.

Through the use of this process, respondents are asked to respond quantitatively and qualitatively to the same subject matter. A comparative analysis should be relatively straightforward and allow researchers to identify areas of similarity and difference, both within the responses of a given participant and among the various participants.

10.6.2 Phenomenological Interviews and Small *n* Randomization

Qualitative interviews conducted using a phenomenological focus prioritize the interviewee's interpretation of events. This qualitative methodology lends itself particularly well to exploration of clients in substance abuse treatment, in addition to bridging the gap between practitioners and academics referred to in the first section. Therapists and counselors are trained to take a client-centered approach and to elicit information that contributes to the understanding of the client's perceptions; this suggests that these individuals are phenomenologically oriented, so, even if they lack prior qualitative analysis experience, they may acquire the necessary skills with minimal training.

Phenomenological approaches to qualitative inquiry come in a few "flavors," including hermenuetic phenomenology, reflexive phenomenology, and interpretative phenomenological analysis. Finlay (2011) provides comprehensive information on the various types in a text written especially for therapists.

Commonalities among phenomenological approaches include identification of the researchers' or therapists' prior beliefs, use of semi- or unstructured interviews to focus on how clients perceive and make meaning of the experience or phenomena of interest, lengthy and sometimes multiple interviews, and, when multiple clients are interviewed, an analytic process directed toward identifying the commonalities in their descriptions of the experience.

The small *n* or single-case randomization techniques described by Dugard et al. (2012) provide an ideal complement to this phenomenological approach to qualitative interviewing. This process allows statistical hypothesis testing for treatment effects to be performed on a single or small number of individuals, by using data gathered at randomly chosen assessment or observation points. Bulté and Onghena (2008) developed and describe a software package that facilitates single-case randomization designs through the use of open-access software program R (R Core Team 2013). As with the factorial survey described above, statistical data, in this instance from randomization tests, can be compared to client's perceptions and gleaned from analysis of phenomenological interviews.

10.6.3 Alternative Data Types

Our third recommendation is based on sources of data rather than design. Qualitative researchers may be more inclined than quantitative researchers to explore alternative types of data, such as visual or literary arts, audio and video recordings, and other expressive forms. Technology, in particular, has had an impact on the evolution of accessibility to expressive media. In the earliest of photo elicitation, researchers passed out disposable cameras so participants could shoot film pictures; the increasing prevalence of sophisticated mobile phones and other devices has resulted in the ability of many individuals throughout the world to create digital photographs and videos that they can quickly and easily disseminate via wireless internet access.

Both qualitative and quantitative researchers have embraced some elements of technology; online surveys are increasingly the distribution method of choice in some settings, while qualitative researchers have explored email interviews and virtual ethnography of Internet "communities." While there is tremendous potential for multimedia enhanced or informed mixed methods research designs, it is wise to caution that not all institutional review boards view online information in the same way. As such, it is advisable that researchers discuss their plans in explicit detail before soliciting or mining Internet data. Whether or not review boards consider Internet research exempt, researchers are encouraged to be open and honest when participating in virtual communities, as well as to thoroughly de-identify any data.

Following is a limited list of potential mixed, qualitative or quantitative methods that reflect current technology. It is the authors' hope that these

provide inspiration to help researchers develop creative research designs incorporating one or more technology or media-based components.

- Interpretive or frequency-based content analysis of Internet interest forums (qualitative or quantitative).
- Use of photographic and video capabilities on mobile phones and/or audio recordings to record participant experiences (qualitative or quantitative).
- Use of mobile phone contact as triggers to report behaviors or responses (qualitative or quantitative).
- Use of mobile phone apps to prompt, tally, or track behaviors over extended period of time (qualitative or quantitative).
- Participant creation of software or web-based visual arts expressions (primarily qualitative).
- Secondary data analysis of website information (qualitative or quantitative).
- Use of blogs as reflective journals for clients (primarily qualitative).

10.7 Closing Thoughts

There is tremendous potential in mixed methods study designs to increase researcher understanding of substance abuse issues at the individual, group, and program level. The material we presented summarizes several of the key considerations for researchers approaching mixed methods designs, though it is by no means comprehensive; interested readers and researchers will want to consult other sources, including many in the subsequent reference list. Furthermore, this chapter has tried to balance a variety of philosophical approaches to research, although it should be noted that many other authors have specific ideas about paradigms and worldviews. Hopefully, the material in this chapter has achieved the twofold goal of providing a starting place for researchers who would like to begin using mixed methods approaches, as well as encouraging current mixed methods practitioners to explore new or creative approaches to substance abuse research.

References

Auspurg, K., Hinz, T., & Liebig, S. (2009). *Complexity, learning effects, and plausibility of vignettes in factorial surveys.* Bibliotheck der Universität Konstanz. Retrieved from http://kops.ub.uni-konstanz.de/bitstream/handle/urn:nbn:de:bsz:352-150806/Hinz%20etal.pdf?sequence=2

Brener, L., Von Hippel, W., Von Hippel, C., Resnick, I., & Treloar, C. (2010). Perceptions of discriminatory treatment by staff as predictors of drug treatment completion: Utility of a mixed methods approach. *Drug and Alcohol Review, 29,* 491–497.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6* (1), 97–113.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research, 1*(1), 8–22.

Bryman, A. (2012). *Social Research Methods* (4th ed.). London, UK: Oxford University Press.

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods, 40*(2), 467–478.

Caelli, K., Ray, L., & Millk J. (2003). 'Clear as mud': Toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods, 2*(2). Article 1. Retrieved from http://www.ualberta.ca/~iiqm/backissues/pdf/caellietal.pdf

Creswell, J. W., Klasse, A. C., Plano Clark, V. L., & Smith, K. C. (2011). *Best practices for mixed methods research in the health sciences.* National Institutes of Health, Office of Behavioral and Social Sciences Research. Retrieved from https://obssr-archive.od.nih.gov/mixed_methods_research/

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). USA: Thousand Oaks Sage.

Dalgaard, P. (2008). *Introductory statistics with R* (2nd ed.). New York: Springer.

Dugard, P., File, P., & Todman, J. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests* (2nd ed.). New York, NY: Routledge.

Dulmer, H. (2007). Experimental plans in factorial surveys: Random or quota designs. *Sociological Methods & Research, 35*(3), 382–409. doi:10.1177/0049124106292367.

Finlay, L. (2011). *Phenomenology for therapists: Researching the lived world.* UK: Wiley.

Ganong, L., & Coleman, M. (2005). Measuring intergenerational obligations. *Journal of Marriage and Family, 67*(4), 1003–1011.

Ganong, L., & Coleman, M. (2006). Multiple segment factorial vignette design. *Journal of Marriage and Family, 68*(2) 455–468.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory.* San Francisco, CA: Sociology.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick: Aldine.

Gorard, S. (2010). Research design, as independent of methods. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral science* (2nd ed., pp. 237–251). USA: Thousand Oaks Sage.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research. Conducting educational research*. UK: Open University.

Green, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed methods research. *Educational Evaluation and Policy Analysis, 11*(3), 255–274.

Green, J. C., & Hall, J. N. (2010). Dialectics and pragmatism: Being of consequence. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral science* (2nd ed., pp. 119–144). USA: Thousand Oaks Sage.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River: Prentice-Hall.

Hesse-Biber, S. N. (2010). *Mixed methods research: merging theory with practice*. New York: Guilford.

Hox, J. J., Kreft, I. G. G., & Hermkens, P. L. J. (1991). The analysis of factorial surveys. *Sociological Methods & Research, 19*(4), 493–510.

Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research, 34*(3), 334–423.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112–133.

Liebow, E. (1967). *Tally's Corner: A study of Negro Streetcorner Men*. Boston: Little, Brown, & Company.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook* (3rd ed.). USA: Thousand Oaks Sage.

Morse, J. M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek: Left Coast.

National Institutes of Health (NIH). (2016). *NIH RePORTER.* Retrieved October from https://projectreporter.nih.gov/

Patton, M. Q. (1997). *Utilization focused evaluation* (3rd ed.). USA: Thousand Oaks Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). USA: Thousand Oaks Sage.

Rayburn, R. L. (2013). Understanding homelessness, mental health and substance abuse through a mixed-methods longitudinal approach. *Health Sociology Review, 22*(4), 389–399.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15–67). Beverly Hills, CA: Sage.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). USA: Thousand Oaks Sage.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health, 23,* 334–340.

Sandelowski, M., & Barroso, J. (2003). Classifying the findings in qualitative studies. *Qualitative Health Research, 13*(7), 905–923.

Spradley, J. P. (1970). *You owe yourself a Drunk: An ethnography of Urban Nomads*. Boston: Little, Brown, & Company.

Stack, C. B. (1975). *All Our Kin*. New York: Basic books.

Strauss, A., & Corbin, J. (2008). *Basics of qualitative research* (3rd ed.). USA: Thousand Oaks Sage.

Tabachnick, & Fidell, L. (2012). *Using multivariate statistics* (6th ed.,). Upper Saddle River, NJ: Pearson.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. USA: Thousand Oaks Sage.

VanDeMark, N. R. (2007). Policy on reintegration of women with histories of substance abuse: A mixed methods study of predictors of relapse and facilitators of recovery. *Substance Abuse Treatment, Prevention, and Policy, 2*(1), 28.

Wagner, K. D., Davidson, P. J., Pollini, R. A., Strathdee, S. A., Washburn, R., & Palinkas, L. A. (2012). Reconciling incongruous qualitative and quantitative findings in mixed methods research: Exemplars from research with drug using populations. *International Journal of Drug Policy, 23,* 54–61.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38,* 505–520.

Wallander, L., & Blomqvist, J. (2009). Modeling ideal treatment recommendations: A factorial survey of Swedish social workers' ideal recommendations of inpatient or outpatient treatment for problem substance users. *Journal of Social Service Research, 35,* 47–64.

Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th ed.). Thousand Oaks, CA: Sage.

Part IV
Measurement Issues

Assessment of Substance Misuse

11

Timothy J. Grigsby, Steve Sussman, Chih-Ping Chou, and Susan L. Ames

11.1 Introduction

Assessment of drug use has traditionally consisted of gathering basic information on past drug use behaviors (e.g., type, frequency, or consistency), as well as hypothesized risk and protective factors for the initiation or regular use of tobacco, alcohol, and other drugs. Such measurement is typically completed for research purposes and to detect changing trends that may flag potential future problems (e.g., see http://www.monitoringthefuture.org/; accessed 8-28-2016). However, there are several clinical and theoretical circumstances where assessing individuals for problematic drug use (a proxy measure of drug misuse) is needed. Here, we describe the appraisal of psychological, physical (biological), interpersonal, and socio-environmental consequences resulting from excessive consumption or intoxication of alcohol or other drugs to form a definition of drug misuse.[1]

Research on drug misuse has broad implications for the prevention and treatment of alcohol and drug use disorders. For example, identifying common life consequences among disparate patterns of drug use can assist researchers in developing interventions and treatment protocols in order to maximize effectiveness across the spectrum of drug use disorders. Similarly, the identification of related consequences among users of different drugs of abuse can aid social workers, public officials, and concerned family members in the early identification of drug misuse with the goal of improving the prognosis for the drug user. Clinical implications of drug misuse assessment may include further testing to determine if a formal diagnosis of a drug use disorder is applicable, determination of need for inpatient or outpatient chemical dependence treatment, and means of potential resolution for drug-related negative consequences. The initial goal of assessment is to determine the nature of an individual's involvement with drugs and to assess their psychological and medical status, psychosocial functioning, social support networks, attitudes toward drug use, previous quit attempts, and motivation for future cessation attempts.

T.J. Grigsby (✉)
Department of Kinesiology, Health, and Nutrition, University of Texas, San Antonio, San Antonio, TX 78249, USA
e-mail: timothy.grigsby@utsa.edu

S. Sussman · C.-P. Chou,
Institute for Health Promotion and Disease Prevention Research, University of Southern California, Los Angeles, CA, USA
e-mail: ssussma@usc.edu

C.-P. Chou,
e-mail: cchou@usc.edu

S.L. Ames
School of Community and Global Health, Claremont Graduate University, Claremont, CA, USA
e-mail: susan.ames@cgu.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_11

[1]We use the term drug misuse, as defined above, in place of abuse as the latter can refer to a chronic pattern of drug misuse behaviors that—when considered in combination with issues of chemical dependence—form a diagnosable drug use disorder (APA 2013). The goal of this chapter is to review and critique assessments that appraise proximal life consequences that result from drug use behavior.

There are two general types of assessment professionals employ to detect potential drug misuse. First, assessment may be used as for screening ("proactive assessment"). As examples, professionals may be asked to screen for drug misuse among employment candidates for jobs in which public safety is paramount (e.g., lifeguards, vehicle drivers, babysitters, or airplane pilots), or those participating in athletics (e.g., to detect presence of performance-enhancing drugs). Second, assessment may be used as a tool to clarify a presenting medical or socio-psychological problem ("reactive assessment"). As examples, professionals may be asked to assess those who suffer from certain physical or medical problems which may be related to drug misuse (e.g., liver problems) or those who have been garnering complaints or concern from significant others or public authorities (e.g., because the person being assessed has been showing behavioral signs of drug misuse such as slurring words, shaking, showing uncontrolled anger; showing physical signs of drug misuse such as alcohol-breath, clothing odors, sweating uncontrollably, poor posture, or red eyes); or performance-related signs of drug misuse such as demonstrating lackluster performance at work or home, or experiencing legal difficulties such as DUI or drunken and disorderly behavior. Both proactive and reactive assessments of drug misuse are important to differentiate drug misuse from other problems, and to facilitate appropriate treatment planning, so as to begin the process of helping an individual arrest negative consequences and permit recovery of functioning, as well as minimize consequences to significant others and protect the public.

There are numerous techniques available for the assessment of drug misuse that can be organized by approach (e.g., interview, self-report, biological). This chapter reviews frequently used assessments of drug misuse, provides the strengths and limitations of each respective approach, discusses issues related to secondary data analysis and field research, and concludes with recommendations for transdisciplinary efforts in drug misuse assessment. The assessment approaches reviewed in this chapter include examples of common self-reported drug use assessments, unstructured examinations and interviews (e.g., mental status, physical, and psychiatric screening, general drug use history, current general behavioral and family characteristics, corroborative reports), structured interviews (e.g., life and drug use history interviews), structured ("brief") inventories of alcohol and drug use consequences, comprehensive drug use and psychological inventories, and biochemical measures. Both unstructured and structured assessments can be used in clinical and research settings, although some examples may be more applicable to one or the other setting.

11.2 Self-Reported Frequency, Quantity, Method of Drug Use, and Family History of Drug Use

Assessing the frequency and quantity of drug use may not be essential in assessing the degree of drug misuse; however, it is nonetheless associated with drug-abuse-related dysfunction (Rychtarik et al. 1998, 1999). For example, there are some individuals who experience severe consequences while using relatively low levels of drugs (e.g., flushing and palpitations that some Asian groups experience with alcohol use). Recently, Grigsby et al. (2014) found that relatively low frequency drug use among Hispanic adolescents who experienced a high number of drug use consequences during adolescence was associated with depression in early adulthood. Conversely, there are some individuals who appear to experience few consequences despite relatively high levels of regular drug use. However, high quantities of intake are more commonly correlated with occupational, educational, social, and medical impairment (Sussman and Ames 2008).

Frequency of drug use indicates how often individuals are using a drug. Frequency of use can

be measured through self-reported estimates of lifetime use, yearly use, monthly use, weekly use, and/or daily estimates of use. A prominent measure of drug use frequency is the timeline follow-back method (Sobell and Sobell 1992). This technique requires individuals to recall previous drug use behavior over a specified time period using memorable life events and a personalized calendar to assist with recall, and has been shown to be a reliable and valid self-report measure of alcohol and illicit drug use in the general population (Sobell and Sobell 1992; Hjorthøj et al. 2012; Robinson et al. 2012). Recall times are free to vary, but recent evidence suggests that one-week recall periods are more efficient for capturing atypical substance use patterns whereas a recall time longer than 2 weeks is more appropriate for estimating stable, consistent substance use behavior (Buu et al. 2014).

Unfortunately, relying on frequency of drug use as the primary assessment of drug misuse lacks precision, as it is not designed to compare disparate patterns of drug use. For example, consuming one "hit" of methamphetamine can lead to a psychological effect of feeling high for 6–12 h (Krasnova and Cadet 2009), whereas generally it would take several servings of standard alcoholic beverages to achieve a similar sense of drug-induced euphoria. As such, recent frequency of drug use may differ somewhat due to psychoactive effects of the drug—and not entirely due to craving. Of course, regular use over longer periods of time (e.g., a year or longer) is commonly accepted as an indicator of high use frequency.

More often than not, descriptions of recent use does not indicate the length of time or the extent of possible drug misuse, but it does help to disclose the most current and reliable autobiographical events. Quantity of use is more predictive of problems or disruptive drug use and is used to describe problem behaviors such as binge drinking or heavy drinking (Newcomb and Felix-Ortiz 1992). The National Household Survey on Drug Abuse (http://www.samhsa.gov/data/population-data-nsduh) among other national surveys, defines binge drinking as consuming five or more drinks on one occasion for at least 1 day in the past 30 days. More recently, the National Institute on Alcohol Abuse and Alcoholism (2004) created gender-specific definitions of binge drinking where consuming four or more drinks within approximately two hours for women and five or more drinks within approximately two hours for men constituted binge drinking. The evolution in the definition of binge drinking has led to an increase in the identification of women drinking at potentially dangerous levels (Chavez et al. 2011). Alternatively, heavy drinking is defined as drinking five or more drinks on the same occasion on 5 or more of the past 30 days (Substance Abuse and Mental Health Services Administration 2002). The use of a drinking diary can help improve the quality of self-reported data regarding binge or heavy drinking (e.g., Dennis et al. 2004); however, compliance with keeping careful records of intake can be problematic (e.g., social desirability confounder, trying to write records while drunk). While the most systematically measured quantity of use has been done in relation to alcohol use, the association of—relatively—high quantities of use for various illicit drugs with negative consequences (e.g., overdoses, fainting, loss of behavioral control) is well known.

Assessing the method of drug intake may also help one to understand the level of misuse for drugs with multiple points of entry into the body. For instance, many individuals with crack/cocaine addiction may have started their use by snorting powdered cocaine. Eventually, they may switch to a different form of the drug—smoking crack that is cheaper and readily available in small quantities—and that immediately potentiates dopamine transmission in the brain. Some drugs, such as heroin, have become increasingly cheaper and more potent in recent years and dependence can easily occur through smoking. Assessment of family history of drug use may provide further information to help researchers and clinicians assess drug availability in the home environment, understand the user's

attitudes toward drug use, perceptions of drug-related problems and consequences, and probability of relapse if treatment is considered.

11.2.1 Strengths and Weaknesses of Using Self-Report Measures of Drug Use

Self-report measures of drug use typically consist of simple questions that provide useful information regarding type, frequency, and method of drug use. This mode of assessment is common among national surveys of drug use behavior in the United States (U.S.) including the National Survey on Drug Use and Health (NSDUH) (Substance Abuse and Mental Health Services Administration 2013). Such information is useful for detecting trends in alcohol and other drug use behavior, but does little to help identify drug misuse at the individual level. More recently, nationwide surveys such as the NSDUH and the National Comorbidity Survey have begun to include items on personal and social consequences of drug use and measures to capture risk and protective factors for drug abuse. Of course, the goal of this assessment has been to estimate the prevalence of drug use disorders by DSM-IV or ICD-10 diagnostic criteria (Kessler et al. 2005a, b; Substance Abuse and Mental Health Services Administration 2013), and as such limits the assessment of drug use consequences to those included in the diagnostic criteria for alcohol and drug use disorders, and oftentimes overlooks the extent to which drug use consequences impact other important domains of functioning.

Self-reported drug use remains the most common mode of assessing drug misuse, but there remain several additional distinct limitations to the use of self-report data for this purpose. First, there are numerous and complex patterns of drug use and related consequences making comparisons across individuals, or groups, difficult. These patterns are a function of drug availability, context of use, sociodemographic features including age, gender, race/ethnicity, SES, biological differences in metabolism and neurochemistry, among others. Second, a lack of sensitivity to other pertinent factors including length of drug use may confound the appraisal of drug misuse and produce results that underestimate the prevalence of misuse in community samples, incorrectly categorize the severity of misuse or lead to insufficient treatment planning and prognosis for the individual. Additionally, there are concerns unique to the disclosure of sensitive information. Self-report data are vulnerable to issues of dishonesty, memory bias, social desirability, and response demand problems (Marlow and Crowne 1961; Choi and Pak 2005). Such data can produce unreliable information for research and clinical purposes especially in cases where the drug user is not inclined to undergo detection, experimental intervention, or treatment. In these instances, the use of corroborative findings may be useful to increase the accuracy of data collection. These might include family members' reports, nondrug using friends' reports, or biochemical assessment methods (described below). Of course, judgments regarding the severity of drug use behavior should be based on variables such as the age of the drug user. For example, any use of an illicit drug, or drugs such as inhalants, by a child or young teen indicates potential immediate danger requiring immediate intervention. As such, self-report measures should be interpreted with proper judgment provided the context of the responses provided.[2]

11.3 Unstructured Presenting/Intake Interviews

Unstructured interviews typically take an approach where there is no standardization of questioning. Instead, the researcher, clinician, or counselor is responsible for determining what questions are—and are not—asked and how the responses can be used to make a clinical diagnosis and for treatment planning (Summerfeldt and Antony 2002). The *intake interview* aims to

[2]Chapter 13 in this volume provides additional discussion on this topic

collect detailed information regarding one's drug use history and related consequences while screening for possible comorbidities and is used primarily in clinical settings.

While this approach does not have a structured format, where one response prompts a subsequent line of questioning, the American Psychiatric Association (APA) (2006) suggests that practitioners follow a general outline consisting of questions in 13 "domains": identifying information, presenting problem/chief complaint, history of the presenting problem, family history, relationship history, developmental history, educational history, work history, medical history, substance use, legal history, and history of previous counseling. Each domain can provide the interviewer with the information needed to make a diagnosis (Jones 2010) and can assist with treatment planning. Sometimes, the interviewee may deny a chief complaint pertaining to drug misuse such as complaints of drug-induced behavior by friends and coworkers. The interviewer may need to ask questions in such a way as to assess the interviewee's perceptions of what others think. Corroborative reports may be quite important in such circumstances.

Relevant questions that might be asked during the substance use assessment of an intake interview include assessments of *drug use behavior* (Have you taken or tried any drugs? What do you use? How much (i.e., quantity and frequency)? How old were you when you first used?), *subjective appraisal of psychological and physiological reactions* (What were your first experiences with drug use like? What are your experiences with drug use like now? How much control do you think you have over your drug use? How long do your using episodes last? What happens? What do the drugs do for you? How do they make you feel?), *alcohol/drug refusal self-efficacy* (In what situations do thoughts about using drugs just "pop" to mind?), *other behavioral problems* (Do you seem to lose control over any other areas of your life? How about gambling? Sex? Spending? Eating? Exercising a lot?), *relevant contextual factors* (Are you studying or working very long hours? Do you experience any suicidal or homicidal ideation? What were you taught about your culture's attitudes toward drug use? Do you live in an area in which a lot of drugs are available? How many liquor stores are near your home? Are the names of the drug dealers well known in your neighborhood? Do your peers use any drugs? Does your best friend use drugs? Do your parents use drugs?), *negative drug use consequences* (Have you experienced any legal or social problems from drug use? Have you ever caused or suffered any property damage while using drugs?), *psychiatric comorbidities* (Have you ever gone to a psychiatrist or other professional for mental health concerns?), *and lifetime traumatic events* (Have you experienced any sexual or physical abuse?) Any disclosure of intended physical harm to self or others would need to be reported by the interviewer to the appropriate agency. These types of questions are best approached with circumspect.

In addition, a *mental status examination* generally is conducted as a systematic means of gathering psychological and behavioral data during the intake interview. The purpose is to provide an initial screening of an individual's mental health status and to help suggest other means of assessment to determine whether a diagnosis of a formal psychiatric diagnosis or comorbid condition should be made. The mental status examination includes the assessment of appearance, attitude and behavior, speech, affect, thought and language, and perceptions and cognitive functioning, such as insight and judgment (Schottenfeld and Pantalon 1994). When performing a mental status examination, the following questions are examples of those that might help to provide a guideline to determine whether an individual is suspected of drug abuse or other psychopathology. Relevant questions to consider during the medical status examination include: Does the individual appear to be withdrawn, socially isolated, undernourished, agitated or depressed, tired, unable to concentrate, indifferent to pleasurable activities, or unkempt in physical appearance? Is the individual hostile or uncooperative, evasive, or defensive and are there any discrepancies in reports of autobiographical events (i.e., lies, missing information)?

Have any delusions or visual or auditory hallucinations been reported? If so, what were the circumstances? Was the individual under the influence of mood-altering drugs at the time? After answering these questions, the individual might be assessed through a more specific (i.e., structured) interview assessment.

11.3.1 Strengths and Weaknesses of Unstructured Drug Misuse Interview Assessments

The use of interviews that elicit information regarding an individual's prior involvement in drug use and treatment programs, psychiatric facilities, self-help support groups (e.g., 12-step programs), or public sanctions (e.g., court, juvenile hall, camps, or community schools) can be quite useful. These data can assist in understanding an individual's level of drug dependence (i.e., where he/she falls on the drug misuse or use disorder continuum), occurrence of other compulsive behaviors and psychiatric difficulties, beliefs and attitudes of their behaviors, and motivation to stop using. Information regarding the longest period of abstinence endured with the help of a structured environment and without the help of a structured environment is also useful. Anecdotally, many individuals will disclose that while in juvenile hall or prison, or while in treatment as inpatients, they can remain abstinent, but when in the community—without some structure—they are unable to remain abstinent.

The unstructured interview format can produce incredibly detailed information that is specific to the individual being questioned. This can assist clinicians in recommending more thorough assessments, and if conducted properly can be efficacious at determining a diagnosis and subsequent treatment plan (Jones 2010) although some are in disagreement as to whether or not unstructured interview formats have any notable advantages (Samet et al. 2007). That is, as these techniques typically do not have a psychometric structure, it is difficult to determine the reliability and validity of unstructured interview formats used to collect drug use and misuse information. The general lack of consistency between assessments in the frequency, type, and content of the questions asked make it possible to overlook important information across cases. These disadvantages make the use of unstructured interview data troublesome for researchers where the use of parametric statistics to make population level inferences requires reliable and valid assessments of behavior that produce consistent results across individuals and samples.

11.4 Structured Interview Assessments

The structured interview format is designed to present participants and patients with predetermined "closed" questions. Closed questions require simple answers. An example of such a question is "How many months have you experienced this drug problem?" or "When did you last use alcohol before operating a motor vehicle?" This format is advantageous for collecting specific information needed for a research project or to make a preliminary diagnosis in order to begin immediate treatment. Several structured interview assessments are currently available for clinicians and researchers and the subset described in the chapter are presented in Table 11.1.

11.4.1 Adult Structured Interview Assessments

The *Comprehensive Drinker Profile* (CDP, Miller and Marlatt 1984) is a structured interview that was originally designed in 1971 to assess alcoholism in male inpatients, but revised to achieve several clinical and research applications. The 88-item CDP collects detailed information on an individual's alcohol consumption history, motivation, behavior, and self-efficacy and has shown acceptable reliability for regular (consistent) drinking patterns, but may not be optimal for assessing episodic or infrequent, but heavy, drinking sessions (Miller et al. 1992;

Table 11.1 Summary of structured interview measures commonly used to assess alcohol and drug misuse

Measure	Author (Year)	Reliability and validity	Cost and training	Length, administration time and additional comments
Comprehensive drinker profile (CDP)	Miller and Marlatt (1984)	Cronbach's α undetermined for quantitative items Inter-rater agreement: unknown Content and criterion validity established	Formal training recommended Public domain—no cost to use Scoring: varies by gender and subscale—consult manual	88-items Administration time: 60–80 min Commonly used at intake to screen for alcoholism Incorporates items from the MAST and has supplemental brief and follow-up versions
Substance dependence severity scale	Miele et al. (2000a, b)	α = 0.69–0.91 (varies by drug and subscale) Test-retest reliability: ICC = 0.41–0.87 (varies by drug and subscale) Criterion and construct validity established	Training required—costs vary by size of group No copyright costs Scoring: varies by substance—consult user's manual and previous work	7–10 screening items 13 "symptom" items 3 subscales Semi-structured interview Administration time: 15–40 min
Addiction severity index (ASI)	McLellan et al. (1980)	α = 0.46–0.93 (varies by composite score) Test-retest reliability: ICC = 0.64–0.86 Split half reliability established Content, criterion and construct validity established	Training required—self training manual included, on-site training available Public domain—no cost to use Scoring: Follow manual instructions to develop composite scores for each domain assessed	200-items, 7 domains Brief versions available Semi-structured interview Administration time: 40–80 min Norms based on those seeking treatment for alcohol or drug use problems
Adolescent drug abuse diagnosis (ADAD)	Friedman and Utada (1989)	Test-retest reliability: r = 0.71–0.95 (by domain) 0.83 −0.96 (severity ratings), 0.91–0.99 (composite scores) Inter-rater agreement: r = 0.85–0.97 (by domain) Content and criterion validity established	Training recommended (in person or by video) Public domain—no cost to use Scoring: Each domain is scored to create composite scores—used in combination they present a comprehensive problem profile	150-items, 9 domains Structured interview Administration time: 40–60 min Short version (83-items) available
Comprehensive addiction severity index for adolescents (CASI-A)	Meyers et al. (1995)	α = 0.48–0.80 Content, criterion and construct validity established	Training required—2 day program for $2000 Copyright, but free for research purposes Scoring: sum scores in each domain to develop composite scores. Evaluation of subjective items are done on an individual item basis	Item length varies 10 modules (domains) Semi-structured interview Administration time: 45–90 min Norms based on gender, age group and ethnicity
Adolescent diagnostic interview (ADI)	Winters and Henly (1993)	Test-retest reliability: Kappa (K) = 0.53–0.78 (alcohol) and 0.52–0.79 (marijuana) Inter-rater agreement: K = 0.66–0.96 (alcohol)	Training required— Copyright costs: $75.00 per kit (manual plus five administration booklets) Scoring: three types of scores can be constructed	213 possible items Structured interview with skip patterns Administration time: 45–90 min

(continued)

Table 11.1 (continued)

Measure	Author (Year)	Reliability and validity	Cost and training	Length, administration time and additional comments
		and 0.82–0.97 (marijuana) Content, criterion and construct validity established	(individual, abuse and dependence symptoms, DSM-III diagnosis). Consult manual for instructions for scoring instructions for each type	Criterion-referenced interview using diagnostic information from DSM-III-R for psychoactive substance use disorders Reliability and validity estimates based on clinical samples
DSM 5 SCID	APA (2013)	Forthcoming	Training required Copyright—cost, availability and scoring procedures forthcoming	Forthcoming

Miller and Del Boca 1994). The CDP includes a section related to alcohol-related problems that can be used to assess drug misuse. The measure produces two scores: a score for alcohol abuse symptomology (problems) derived from the Michigan Alcohol Screening Test (described below) and an indicator of physical dependency. This interview was developed to determine treatment modality. The CDP has been used more extensively in clinical settings, and while reliability and validity have been established, exact estimates are not commonly reported in the literature for this measure. Three additional instruments were designed to complement the CDP including a brief interview [the Brief Drinker Profile (Miller and Marlatt 1984)], a parallel outcome measure for use in treatment settings [the Follow-Up Drinker Profile (Miller and Marlatt 1987)], and a corroborative measure given to "significant" others that can be compared to responses on the CDP (the Collateral Interview Form Miller and Marlatt (1987).

The *Substance Dependence Severity Scale* (SDSS; Miele et al. 2000a) is a 13-item clinician-administered structured interview that was developed to assess severity and frequency of dependence across a range of drugs, based on the DSM-IV diagnostic criteria for substance use disorders. The test/retest, joint rating, and internal consistency reliabilities across alcohol, cocaine, heroin, marijuana, and sedative users ranged from fair to good (Interclass correlation coefficients (ICCs) = 0.41–0.87) with estimates ranging by drug and subscale. In a sample of alcohol, cocaine and heroin users entering substance abuse treatment, the SDSS demonstrated convergent and discriminant validity when compared to other measures of substance abuse (e.g., the Addiction Severity Index) and survival analyses demonstrated predictive validity when time to first substance use was used as an outcome (Miele et al. 2000b).

The *Addiction Severity Index* (*ASI*) is a structured clinical research interview designed to provide information about various areas of an individual's life in which there often exists dysfunction associated with drug abuse. The 200-item ASI assesses problem areas including medical, legal, drug abuse, alcohol abuse, employment, family, and psychiatric problems. Reliability and validity data for the ASI have been extensively reported (McLellan et al. 1980, 1985; Rounsaville et al. 1986, Leonhard et al. 2000; Mäkelä 2004). Early work by McLellan et al. (1985) has resulted in a strategy for obtaining a composite score based on the sum of several individual questions within specific problem areas. However, use of the composite score as a primary measure of drug use and related problems should be done with caution. As outlined by Mäkelä (2004), the composite scores are not always independent of one another, and

could be an indicator of poor reliability instead of discriminant validity. Second, using composite scores for longitudinal analysis may be troublesome as changes in single items could be interpreted as changes in global functioning for a particular problem area. This may be the result of combining subjective evaluation items (i.e., how severe the participant views their alcohol/drug use problems) with objective evaluation items (i.e., how many days in the past 30 days the participant has used alcohol or other drugs) (Wertz et al. 1995). Previous work has demonstrated that self-administered formats of the ASI produce similar results to the interview format. Rosen et al. (2000) found correlations for composite scores ranging between 0.59 and 0.87 for legal, drug abuse, alcohol abuse, employment, family, and psychiatric problems. Correlations for medical problems were slightly lower ($r = 0.47$) and participants acknowledged more drug use and psychiatric symptoms in the self-report format.

11.4.2 Adolescent Structured Interview Assessments

The *Adolescent Drug Abuse Diagnosis* (*ADAD*) is a comprehensive structured interview consisting of 150 items used to assess substance abuse and other problem areas of which 83 items can be used to assess change following treatment for substance abuse. The format is adapted from the adult tool, the Addiction Severity Index (ASI, McLellan et al. 1980). This interview produces composite scores—using a formula to weight selected item scores in each domain—rating the severity of problems in nine life areas including: medical, school, work, social relations, family relationships, legal, psychological, and alcohol and drug use (Friedman and Utada 1989). Responses are gathered on a ten-point severity scale, but three problem checklists (medical, school, and family) use a simple yes/no format to shorten interview time. The measure has demonstrated good test–retest reliability and inter-rater agreement for all subscales.

The *Comprehensive Addiction Severity Index for Adolescents* (*CASI-A*) is an instrument designed to provide an in-depth, comprehensive assessment of the severity of adolescents' addiction and problem consequences. This structured interview is also adapted from the Addiction Severity Index (ASI; McLellan et al. 1980) with the item length varying based on built-in skip patterns to avoid unnecessary follow-up questions. Ten domains of functioning are assessed and include the following: psychological, peer relationships, family history, sexual relationships, physical abuse, significant life changes, use of free time, substance use effects and treatment experiences, leisure activities, educational experiences and plans, legal history, and psychiatric status, including prior treatment experiences (Meyers et al. 1995). The scale has continued to be refined and has been revised as the *Comprehensive Adolescent Severity Inventory* (*CASI*) modeling the criteria of the Diagnostic Interview Schedule for Children (DISC-IV) for assessing DSM-IV substance use disorder criteria (Meyers et al. 2006). The CASI has demonstrated excellent internal consistency (Cronbach's alphas exceeding 0.70 for all dimensions of the scale), concurrent validity (as compared to DISC-IV assessments) and predictive validity for drug use at one-month post assessment. Yet, the scale was unable to adequately predict drug use at the six-month assessment (Meyers et al. 2006).

The *Adolescent Diagnostic Interview* (ADI, Winters and Henly 1993) is a 15 min evaluation used to assess the need for treatment of drug misuse among adolescents. This interview was developed based on DSM-III criteria for substance use disorders and evaluates cognitive impairment, psychosocial stressors, interpersonal, and school functioning factors that may contribute to alcohol or drug misuse. The instrument consists of 24 items and has shown good inter-rater and test–retest reliability in both clinical and correctional adolescent samples.

The *Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association* (*DSM*) *Structured Clinical Interview*

is a widely used in diagnosing substance use disorder, and is currently in its fifth iteration (APA 2013). The previous version of this manual, the DSM-IV, contains specific criteria sets for substance abuse, dependence, intoxication, and withdrawal applicable across different classes of drugs (APA 2000). The structured clinical interview was a primary measure of substance abuse and substance dependence disorders, the former pertaining to the exhibition of one or more of the following symptoms due to recurrent use, last 12 months: (a) failure to fulfill role obligations, (b) hazardous use (physical danger), (c) legal problems, and (d) social problems. Substance dependence disorder was intended to be a more severe disorder with three or more of the following symptoms, due to recurrent use, in the last 12 months including (a) tolerance, (b) withdrawal (c) Using more than intended, (d) desiring, but being unable to quit or cut down, (e) taking up a lot of time, (f) other activities being neglected or given up, and (g) continued use despite related psychological or physical problems due to use.

The DSM-V version describes a single "substance use disorder", which is described along a continuum of severity including moderate and severe categories depending on whether one exhibits two to three or four or more of the following symptoms, respectively, due to recurrent use over the last 12 months: (a) using more than intended, (b) desiring, but being unable to quit or cut down, (c) taking up a lot of time, (d) exhibiting craving, a strong desire to use, (e) failure to fulfill role obligations, (f) continued use despite related social problems, (g) other activities being neglected or given up, (h) hazardous use (physical danger), (i) continued use despite related psychological or physical problems due to use, (j) tolerance, and (k) withdrawal symptoms.

This revision combines the criteria of substance abuse and substance dependence from the previous version to create an overarching diagnosis for substance use disorders. The "problems with law enforcement" symptom has been removed from the list due to cultural factors that make it difficult to apply internationally according to the APA. The symptom, "exhibiting craving, a strong desire to use" has been added as a criterion in order to increase consistency with the International Classification of Diseases (10th edition) criteria and indirect evidence and rationale that it is central to the diagnosis and treatment of substance use disorder (Hasin et al. 2013).

The *Structured Clinical Interview for the Diagnostic Statistical Manual* (*SCID*) is a broad-spectrum instrument that adheres to the DSM-V decision trees for psychiatric diagnosis and encourages multiple paths of exploration, clarification, and clinical decision-making, with specific clarification regarding efforts to decrease or control use, continued use despite problems, specific withdrawal symptoms of a drug, and assessment of comorbidity. This interview is a primary measure of substance use disorders in the field of clinical psychology, but requires extensive training to use efficiently. The DSM-V SCID is currently under development and slated for release in the 2014 calendar year. Previous iterations of this instrument is particularly useful when assessing substance misuse in populations with suspected, or known, co-occurring psychiatric problems.

11.4.3 Strengths and Weaknesses of Using Structured Interviews and Assessments

There are several advantages to using a structured interview process as opposed to an unstructured one. First, structured interviews are more effective in collecting quality information, as there are fewer digressions into areas of little substantive concern. This improves the efficiency of the interview process by saving time and effort for the interviewer and interviewee. Second, structured interviews produce consistent results across individuals allowing researchers and clinicians to identify similarities across different drug use experiences. Finally, using this method

of assessment, the researcher can produce consistent results across cases that produce an objective and information-based result.

While this approach may appear advantageous to unstructured interview methods, there are several limitations that should be considered. First, the effectiveness of the structured interview process is a direct reflection of the interviewer's skills. Formal training in structured interviewing is consistent across institutions that produce clinical psychologists, but experience, intuition and rapport with clientele can greatly influence the nature and direction of the clinical interview. Second, structured interviews are also time consuming and expensive to develop leaving little room for error in the development or administration. Third, the constraints on the questions asked and method of evaluation leaves little room for the researcher to explore other pertinent areas for the individual in question. This could lead to oversight of important facts that contribute to drug misuse. Finally, researchers may be dissuaded from using data from structured interviews as the information collected is used to produce a formal diagnosis of a drug use disorder (abuse, addiction, dependency) and not to identify patterns of drug misuse. However, this data could prove useful for researchers to identify patterns of subclinical drug use that may qualify individuals as drug misusers. Such research can produce useful findings to improve our understanding of the transition from regular use to abuse with an emphasis on defining drug misuse with increased precision by accounting for factors such as frequency and length of use, patterns of drug use and need or desire for intervention.

11.5 Structured "Brief" Inventories of Alcohol and Drug Use Consequences (Screeners and Questionnaires)

Numerous self-administered screeners and questionnaires exist that assess the proximal and deleterious short-term consequences of drug use and can be used by researchers to assess the severity of drug use problems in different domains of functioning (e.g., physical, psychological, interpersonal, etc.) or by clinicians to quickly verify whether or not someone being seen due to suspicion of a drug misuse issue is a candidate for further assessment using a structured interview or comprehensive inventory. The consequences assessed in these measures can reflect intrapersonal (i.e., needing more of a drug than desired) or interpersonal (e.g., neglecting caretaking responsibilities, getting into fights with others) struggles an individual encounters while under the influence of drugs. While this section presents and reviews several "brief" measures (Table 11.2), the reader should be reminded that this chapter does not provide an exhaustive list of available measures from every approach. Many measures exist that are designed to assess drug misuse with specific drugs or polydrug use, and are worthwhile of exploration (see Marijuana Problem Scale, Stephens et al. 2000; Young Adult Alcohol Consequences Scale (YAACQ), Read et al. 2006; Risks and Consequences Questionnaire (RCQ), Stein et al. 2010; Marijuana Consequences Questionnaire (MACQ), Simons et al. 2012). There are "brief" versions of some comprehensive assessments discussed earlier that also warrant mention, such as the Brief Drinker Profile (Miller and Marlatt 1984, 1987)—a short version of the Comprehensive Drinker Profile, for example. The measures presented in this section represent a selection of measures used primarily in clinical settings (i.e., CAGE, RAFFT/CRAFFT, MAST/SMAST, DAP, DAST, and AUDIT) followed by measures more commonly adapted for research with clinical and nonclinical populations of drug users (i.e., RAPI, YAAPST, DrInC, InDUC).

The *CAGE* questionnaire (Ewing 1984) is a self-report screening instrument that uses the mnemonic CAGE to assess problems with alcohol. It is a relatively sensitive four-item instrument that assesses attempts to *Cut down* on drinking, *Annoyance* with criticisms of drinking, *Guilt* feelings about drinking and use of alcohol as a morning *Eye-opener*. When a participant responds "yes" to two or more questions, that individual is suspected of having alcohol

Table 11.2 Summary of structured "brief" inventory measures commonly used to assess alcohol and drug misuse

Measure	Author/s (Year)	Psychometrics	Cost, training, scoring	Length, administration time and additional comments
CAGE	Ewing (1984)	$\alpha = 0.69$ Test-retest reliability: $r = 0.80–0.95$ Criterion and concurrent validity established Average ability to detect alcohol abuse or dependence in clinical populations: Sensitivity = 0.71 Specificity = 0.90	No training necessary Public domain—no cost to use Scoring: sum items for range of 0–4 where higher scores indicate greater severity of alcohol use problems	4-items Administration time: under 1 min Can be used with adolescent or adult populations
CRAFFT	Knight et al. (2002)	$\alpha = 0.68$ Criterion and concurrent validity established. Sensitivity and specificity estimates for any problem (0.76 and 0.94), any disorder (0.80 and 0.86) or any substance use dependence (0.92 and 0.80)	No training necessary Copyright, no cost to use Scoring: sum items for range of 0–6 where higher scores indicate greater severity of alcohol and drug use problems	6-items Administration time: 1–2 min Developmentally tailored for adolescent populations. Identifies alcohol and drug use problems
Michigan alcohol screening test (MAST)	Selzer (1971)	$\alpha = 0.83–0.93$ Test-retest reliability: $r = 0.84–0.96$ Content and criterion validity established	No training necessary Cost for copies, no fee for use Scoring: sum items. Score between 0 and 3 indicates no apparent problem, 4 indicates early/middle problem drinker, 5+ indicates problem drinker (alcoholism)	25-items Administration time: 8–10 min Shorter versions available with 9, 10 or 13 items
Drug abuse screening test (DAST)	Skinner (1982)	$\alpha = 0.86–0.94$ Test-retest reliability: $r = 0.71–0.89$ Content, construct and criterion validity established	No training required Copyright: no fee for use Scoring: sum items for range of 0–28 where higher scores indicate more severe problems	28-items 10 and 20 item versions available Administration time: 7–10 min Parallel form to the MAST Useful for detecting problem behaviors for drugs other than alcohol
Drug and alcohol problem (DAP) quickscreen	Schwartz and Wirtz (1990)	$\alpha = 0.46$ Further indicators of reliability and validity are unknown	No training required Public domain (no cost) Scoring: sum items for range of 0–30 where higher scores indicate more severe problems	30-items Administration time: 7–10 min Primarily used as a screening tool by physicians

(continued)

Table 11.2 (continued)

Measure	Author/s (Year)	Psychometrics	Cost, training, scoring	Length, administration time and additional comments
Alcohol use disorders identification test (*AUDIT*)	Saunders et al. (1993) and WHO (1998)	α = 0.75–0.94 Test-retest reliability: r = 0.64–0.93 Content, construct and criterion validity established	Recommended Training module: $75.00 Copyright: no fee for use Scoring: sum items for range from 0 to 40 where higher scores indicate more severe problems	10-items, 3 subscales Administration time: 2–5 min Norms based on heavy drinkers and alcoholics Brochure available to aid with scoring, interpretation and intervention approach
Rutgers Alcohol Problem Index (RAPI)	White and Labouvie (1989)	α = 0.92 Test-retest reliability: r = 0.83–0.92 Criterion validity established	No training required Copyright: no fee for use Scoring: sum responses to create a continuous variable indicating the frequency of alcohol-related problems for a specified time period	18-items Original version (23-items) Administration time: 10–14 min Norms based on community adolescent sample and inpatient (drug and alcohol problem) sample Validated for use with polysubstance users
Young adult alcohol problems screening test (YAAPST)	Hurlbut and Sher (1992)	α = 0.83 Test-retest reliability: r = 0.73 (1 year interval) Content, criterion and construct validity established	No training required Copyright—no fee to use Scoring: Responses are coded yes (1) and no (0), and summed for a range of 0–27 where higher scores reflect more problem use	27-items Administration time: 5–7 min. Used to assess alcohol problems at lifetime, past year and specific time intervals
Drinker inventory of consequences (DrInC)	Miller et al. (1995)	α = 0.70–0.94 (by subscale for past 3 month consequences) Test-retest reliability: r = 0.79–0.96 (by subscale and time interval) Construct validity established	No training required Copyright—no fee to use Scoring: sum total scale and each subscale for comparison with gender based profiles forms. Higher scores indicate higher levels of alcohol use consequences	50-items, 5 subscales Brief version available Administration time: 10–15 min Norms based on inpatient and outpatient clients seeking treatment for alcohol use disorders Factor structure unstable across populations and time points

(continued)

Table 11.2 (continued)

Measure	Author/s (Year)	Psychometrics	Cost, training, scoring	Length, administration time and additional comments
Inventory of drug use consequences (InDUC)	Tonigan and Miller (2002) and Blanchard et al. (2003)	$\alpha = 0.96$ (estimate excludes 5-item control subscale) Test-retest reliability: $r = 0.34$–0.93 for specific scales using lifetime estimates Construct validity established	No training required Copyright—no fee to use Scoring: sum total scale and each subscale where higher scores indicate higher levels of drug use consequences	50-items, 5 subscales Administration time: 10–15 min Alternative form to DrInC for general drug use consequences Recent (InDUC-2R) and lifetime (InDUC-2L) versions of scale available Scores for intrapersonal consequences are not stable over time

problems. These questions can be adapted for other drug use, as well, by replacing the word *drinking* with *drug use*, and *a morning eye opener* with *the drug to get you started in the morning*. The focus of this questionnaire is on consequences of use related to an individual's response to others' perceptions of his or her use, resultant feelings, and attempts to quit. Attempts to change behavior may or may not come from outside sources (e.g., one's limited social group) that, in turn, may cause guilt feelings or cognitively based conflict. The test has demonstrated marginally acceptable reliability ($\alpha = 0.69$), but has shown adequate levels of sensitivity and specificity to detect alcohol abuse or dependence.

The *RAFFT* (*r*elax, *a*lone, *f*riends, *f*amily, *t*rouble) *test* was developed similarly to the CAGE but as a brief screen for teenagers. It can be applied to emerging adults as well. The RAFFT consists of five items (e.g., "Do you drink to relax, to feel better about yourself, or to fit in?") that were not originally intended to work as a validated scale of drug misuse (Riggs et al. 1989). There has been a dearth of psychometric validation with the RAFFT measure. However, Knight et al. (1999) adapted several questions from the RAFFT, RAPI, and DAP to create a brief screening of alcohol and other drug abuse resulting in the nine-item CRAFFT test—which was later reduced to six-items (*c*ar, *r*elax, *a*lone, *f*orget, *f*amily or friends complain, *t*rouble). Items address riding in a car driven by someone under the influence, drinking or using to relax, drinking or using alone, forgetting things while drinking or using, family or friends telling one to cut down, and getting into trouble while under the influence. The CRAFFT has demonstrated good convergent validity and nearly acceptable internal consistency ($\alpha = 0.68$). The items on this assessment represent neurobiologically based (drinking to relax), cognitively based (poor decision-making, as in riding in a car driven by someone under the influence, forgetting things while drinking or using), and socially based (drinking alone) drug use motivations, as well as interpersonal-based consequences of use (family or friends telling one to cut down) and larger social consequences (getting into trouble while under the influence). Moreover, the CRAFFT is capable of detecting problems related to alcohol and other drug use whereas most similar measures capture alcohol only (Knight et al. 2002).

The *Michigan Alcohol Screening Test* (*MAST*) is a 25-item questionnaire used to screen for consequences of problematic alcohol use and perceptions of alcohol-related problems. This

questionnaire was originally developed to place drinkers into early (mild impairment), middle (moderate impairment), and late (severe impairments) stages (or levels of impairment) of alcoholism (Selzer 1971). This measure can be self-administered and used to identify abnormal drinking by addressing social and behavioral consequences (Selzer et al. 1975). More than a decade after its introduction, a review of studies using the MAST concluded that the scale had acceptable inter-item consistency (Cronbach's $\alpha = 0.83–0.93$), but tended to over diagnose alcoholism (Gibbs 1983). The SMAST is a shorter ten-item version that is relatively effective in discriminating alcoholics from nonalcoholics. The items are designed to describe extreme drinking behaviors and to establish the presence of negative consequences of excessive alcohol consumption. Examples of discriminating items are as follows: Have you ever attended a meeting of Alcoholics Anonymous? Have you ever gone to anyone for help about your drinking? Have you ever been in a hospital because of drinking? The Drug and Alcohol Problem (DAP) Quickscreen consists of 30 yes/no items and discriminates well between high-risk and low-risk users (Schwartz and Wirtz 1990). This assessment was developed for use by pediatricians to assess for adolescent alcohol and other drug abuse and includes the prototypical item: "Has anyone (friend, parent, teacher, or counselor) ever told you that they believe that you may have a drinking or drug problem?" Schwartz and Wirtz (1990) concluded that four items accounted for 70% of the variation between high-risk and low-risk users. Further analyses with these four items (Knight et al. 1999) showed poor internal consistency ($\alpha = 0.46$), which is to be expected, as it was not intended as a stand-alone scale.

The *Drug Abuse Screening Test* (DAST; Skinner 1982) can be self-administered as a screener of consequences and problem experiences from drug abuse and dependence. The scale has demonstrated good internal consistency ($\alpha = 0.86–0.94$). Additionally, research has found the ten-item version (DAST-10) to have comparable reliability in addition to strong criterion validity ($r = 0.31–0.39$) and construct validity ($r = 0.40$; Yudko et al. 2007).

The *Alcohol Use Disorders Identification Test* (*AUDIT*), developed by the World Health Organization, is an index of consequences and problems experienced from drinking over the past year (Saunders et al. 1993). The AUDIT can be administered by a clinician orally at initial screening, or as a self-report measure for researchers. Three questions assess quantity and frequency of use, 4 questions assess alcohol-related problems, and three questions assess dependence symptoms (alpha range, 0.75–0.94; Babor et al. 1992; Saunders et al. 1993). After summing responses for the ten items, the practitioner can determine a risk level and employ suggested intervention strategy. For example, an individual with an AUDIT score 16–19 would have a "zone III" risk and it would be recommended to intervene with simple advice in addition to brief counseling and continued monitoring. The AUDIT has very good psychometric properties (e.g., sensitivity, specificity, and internal consistency) and the measure has been validated across a wide range of populations (see Allen et al. 1997; Connors and Volk 2003; Reinert and Allen 2007). Test–retest reliabilities, assessed over varied timeframes, yield a median coefficient alpha of 0.83 (range 0.64–0.93).

The original *Rutgers Alcohol Problem Index* (RAPI; White and Labouvie 1989) consists of 23 items that address consequences of alcohol use related to psychological functioning, delinquency, social relations, family, physical problems, and neuropsychological functioning. Shorter versions of the measure are available (Earleywine et al. 2008) and the measure has been considered a reliable and valid estimate of other drug use consequences (Ginzler et al. 2007). In addition to strong internal consistency ($\alpha = 0.92$), the original measure has been found to correlate highly with DSM-III-R criteria for alcohol use disorders ($r = 0.75–0.95$; White and

Labouvie 1989), and recent evidence showed it significantly correlates with DSM-IV abuse and dependence criteria (r = 0.31–0.82; Ginzler et al. 2007).

Hurlbut and Sher (1992) developed the 27-item *Young Adult Alcohol Problems Screening Test* (*YAAPST*) to capture alcohol-related problems in college-aged populations, and Kahler et al. (2005) later developed a briefer version (24-item). Participants are asked to recall past year and lifetime alcohol-related consequences using a yes/no format. While advantageous for rapid data collection, this approach does not allow investigators to detect the frequency of individual consequences. Reliability of the measure is good (Cronbach's alpha = 0.83) and the measure has demonstrated good test–retest reliability (r = 0.73) for assessing past year consequences.

The *Drinker Inventory of Consequences* (*DrInC*) is a self-administered 50-item measure of the adverse consequences of alcohol use that has demonstrated acceptable reliability estimates across the different subscales (α = 0.70–0.94). The briefer Short Index of Problems (SIP) is a 15-item scale that demonstrated weaker, albeit acceptable, internal consistency in the original study (Miller et al. 1995; Forcehimes et al. 2007). Both the DrInC and the SIP are significantly correlated with daily drinking (r = 0.36 and 0.35, ps < 0.001, respectively) supporting it as a valid measure of alcohol behavior (Forcehimes et al. 2007). It is parallel form; the 50-item Inventory of *Drug Use Consequences* (InDUC; Tonigan and Miller 2002; Blanchard et al. 2003) was designed as a standardized measure of alcohol and other drug use consequences. The measure includes the same five scales as the DrInC measuring (1) impulse control, (2) social responsibility, (3) physical, (4) interpersonal, and (5) intrapersonal consequences. Similar to other drug use consequence scales, this measure excludes topics related to pathological use, dependence symptoms (i.e., craving) and intent to seek treatment. The InDUC has been shown to effectively measure the severity of drug use consequences over time (α = 0.96; Tonigan and Miller 2002). The Short Index of Problems-Alcohol and Drugs (SIP-AD) is a 15-item brief scale that demonstrated comparable internal consistency to the original scale, good concurrent, and discriminant validity that is also sensitive to detecting change in negative consequences over time.

11.5.1 Strengths and Weaknesses of Using Brief Assessments of Alcohol and Drug Use Consequences

These brief assessments of alcohol and drug use consequences have provided researchers with a rapid assessment of possible drug misuse that can be used as preliminary evidence to conduct more extensive assessments. Additionally, the short length and high correlation to longer and more complex assessment tools make them desirable candidates for researchers to use in longer survey batteries where alcohol and drug misuse is a topic of investigation. In fact, recent work has even demonstrated that the AUDIT and DAST measures are equally, and sometimes superior, at identifying substance dependence compared to lengthier inventories (Saitz et al. 2014). These initial findings present a unique opportunity to extend the use of these measures for research and clinical purposes.

The measurement of negative consequences has focused almost exclusively on alcohol use with scales available for other drugs (such as marijuana) only becoming available in the past decade. No measures exist, to the knowledge of the authors, which aim to capture unique consequences experienced by polydrug users. The limited items included in these measures allows for the possibility of overlooking symptoms or behaviors that may be present. For example, an individual may be experiencing a number of intrapersonal complications as a result of their drug use (i.e., depression, anxiety, suicidal ideation, etc.), but these may not have manifested as overt behavioral dysfunctions. As the majority

of these brief scales are interested in interpersonal consequences of drug use, individuals experiencing a wealth of intrapersonal, but not interpersonal, consequences may be overlooked in the screening process. Second, the cross-sectional assessment of drug use consequences has several limitations. Social consequences such as unemployment, arrest, or social isolation may be scarce if the individual has already experienced these as a result of their drug use behavior prior to the time frame of the assessment (i.e., past 3 months, past 6 months, past year, etc.). Another major weakness of these measures has been the lack of cross-cultural validation studies. Cultural attitudes toward defining drug misuse may vary based on the social acceptance of a drug (for example, smoking tobacco is a traditional and ceremonial practice in Native American populations), and appraising drug use consequences—the focus of these measures—may reflect culturally oriented attitudes and beliefs. For example, it is more acceptable to be somewhat intoxicated in public in places like Denmark where the cultural norms are more acceptable of social drinking than many areas of the U.S. (Grønkjær et al. 2011). As such, these brief assessments should be used with caution when working with culturally heterogeneous populations until more work has been compiled that demonstrates the cross-cultural validity for specific measures. Moreover, these brief scales assessing the type and frequency of alcohol and drug use problems should be considered in light of two important variables: subjective evaluation of events and the experience of "positive" consequences such as relieving stress. Some consequences of drug use are taken more seriously, perhaps more likely resulting in a drug use disorder diagnosis and precipitating the ushering of people into treatment. For example, being arrested for an income-generating crime, such as armed robbery as a result of a drug addiction, is obviously a societal, legal problem —one that is placed on the public record, and one in which the perpetrator is likely to be restrained by agents of the public. Yet, if individuals have experienced no legal consequences related to their use, and no obvious interpersonal problems, they are less likely to become identified as drug misusers or diagnosed as drug abusers. Also, there has been a dearth of literature focusing on the experience of positive versus negative alcohol and drug use consequences. Suppose an individual reports experiencing an improved mood, enhanced social skills and general levels of elation when under the influence of a chosen substance. Their cognitive appraisal of associated negative consequences may be less severe compared to a third party (researcher or clinician) evaluating objective measures of frequency and duration of those alcohol or drug-related negative consequences. Finally, McHugh et al. (2014) have cautioned that many of these measures are written above the recommended reading-grade level (5th or 6th grade), and may not be particularly useful for low literacy populations.

11.6 Comprehensive Assessment Inventories of Alcohol and Drug Misuse

Unlike the structured brief inventories of drug use consequences, the comprehensive assessment aims to collect exhaustive information related to an individual's drug use behavior, possible physical, and psychiatric comorbidities and other pertinent factors including information related to demographic information, living situation, context of drug use, etc. Table 11.3 presents the information for the comprehensive measures described below.

Comprehensive Assessments for Adult Populations

The *Alcohol Use Inventory* (AUI; Horn et al. 1986, Littrell 1991; Rychtarik et al. 1998) is a 228-item multiple-choice self-report inventory. It was systematically developed to measure alcohol problems and has demonstrated strong content, construct, and criterion validity. There are 24 subscales with 17 primary scales characterizing individuals along various dimensions with internal consistency ranging between subscales (α = 0.57–0.88). The dimensions are grouped

Table 11.3 Summary of comprehensive assessment measures commonly used to assess alcohol and drug misuse

Measure	Author (Year)	Reliability and validity	Cost and training	Length, administration time and additional comments
Alcohol use inventory (AUI)	Wanberg et al. (1977)	α = 0.57–0.88 (by subscale) Test-retest reliability: r = 0.54–0.89 Content, construct and criterion validity established	No training required Copyright costs: 10 test bundle: $47.00 50 manual score sheets: $33.50 User's manual: $41.50 AUI manual: $52.00 Computerized scoring available for additional fee Scoring: sum subscale totals and compare to normative references for classification	228 items, 24 subscales Administration time: 35–60 min Good for monitoring change in alcohol use behavior over time
The MacAndrew alcoholism scale-revised	MacAndrew (1965, 1989) and Butcher et al. (1989)	α = 0.45 (females) and 0.56 (males) Test-retest reliability: r = 0.78 (females—1 week) and 0.62 (males—1 week) Criterion and construct validity established	No training required for administration of stand-alone version Copyright costs vary by scoring (manual vs. computerized) and scales included (stand-alone vs. MMPI-2 complete measure) Scoring: sum items for range of 0–49 with recommended cutoff of 28 indicating need for further assessment	49-items Administration time: 8–10 min (stand-alone scale) Embedded in the Minnesota Multiphasic Personality Inventory (MMPI-2), but can be administered separately Reliability and validity estimates are of questionable acceptance
Chemical dependency assessment profile	Davis et al. (1989) and Harrell et al. (1991)	α = 0.78–0.88 Test-retest reliability: r = 0.65 and higher (per subscale—individual correlation coefficients not reported) Content and construct validity established	No training required Copyright costs: $20.00 for bundle of 20 tests Software available for report generation and subscale scores Scoring: sum items in each domain where higher scores indicate greater problem severity	232-items, 11 dimensions Administration time: 30–45 min Limited normative sample for interpretation purposes (n = 86)
Inventory of drinking situations	Annis (1982, 1987)	α = 0.87–0.96 (varies by subscale) Test-retest measures unknown for full scale Content and construct validity established	No training required Copyright costs: User's guide: varies by vendor 30 questionnaires: $16.45 Referred to as "IDTS Alcohol questionnaire" Scoring: sum subscales with higher scores indicating heavier drinking in specific situations	100 items, 8 subscales 42-item brief version available and becoming more commonly used than full scale Administration time: 20–30 min for full scale

(continued)

Table 11.3 (continued)

Measure	Author (Year)	Reliability and validity	Cost and training	Length, administration time and additional comments
Inventory of drug taking situations	Annis et al. (1992)	α = 0.70–0.95 (varies by subscale) Test-retest reliability unknown Content and construct validity established	No training required Copyright costs: User's guide: $34.95 30 questionnaires: $16.45 Computerized version available for fee that includes scoring Scoring: sum subscales with higher scores indicating heavier drug use in specific situations	50 items, 8 subscales Administration time: 10–15 min Norms based on specific gender and age group
Substance abuse subtle screening inventory (SASSI)	Miller and Lazowski (1999)	α = 0.93 Test-retest reliability: r = 0.61–0.66 (varies by time and independent study) Criterion validity established	No training required Copyright costs: $3.00 per copy Computer administration and scoring available for fee Scoring: varies by scale, but typically performed on an additive scale with higher scores indicating greater substance use or other domains of interest	93-items (adult version) 100-items (adolescent version) Administration time: 10–15 min Feldstein and Miller (2007) provide a critical review of the measure's reliability, validity and efficacy
Drug use screening inventory-revised (DUSI)	Kirisci et al. (1995) and Tartar and Kirisci (2000)	KR20 = 0.76 (males—average), 0.72 (females—average) Split half reliability: 0.76 (males) and 0.67 (females) Content, criterion and construct validity established	No training required Copyright costs: $2.00 per questionnaire, price for computerized scoring available from company Scoring: sum lie scale to determine validity of responses, then sum each subscale followed by summing all subscale scores for a "global" estimate of problems	159 items, 11 subscales Administration time: 20–40 min Adolescent and adult versions are homologous making longitudinal comparisons possible Can be used to develop a "ranking" of severity from 0 to 100%
Problem oriented screening instrument for teenagers (POSIT)	Rahdert (1991)	α = 0.45–0.79 (varies by domain and study) Test-retest reliability: K = 0.40–0.75 Content and construct validity established	No training required No copyright, free to use Two scoring systems available—see NIDA documentation for details	139 items, ten domains Administration time: 20–30 min Empirically derived cutoff scores for each domain available to distinguish low, medium and high risk using new scoring system

(continued)

Table 11.3 (continued)

Measure	Author (Year)	Reliability and validity	Cost and training	Length, administration time and additional comments
Personal experience inventory (PEI)	Winters and Henly (1989)	α = 0.83–0.97 (varies by subscale) Inter-rater agreement: r = 0.82 Content, criterion and construct validity established	Training not required, but available to "qualified professional users" defined by the APA ethical standards Copyright costs: $165 per kit (five administrations and includes user's manual and computerized scoring/interpretation) Scoring: computerized—consult user's manual	276 items, 27 subscales Administration time: 60–90 min Adult version available
American drug and alcohol survey (ADAS)	Oetting et al. (1999)	α = 0.73–0.96 (varies by study) Content and construct validity established	No training required, measure comes with instructions Costs: varies by order, survey: $0.80–$1.10, score report: $75–$200 Scoring: performed by developer (RMBSI, Inc.)	57 items Administration time: 30–50 min Different versions available depending on age of population (Children's version for students in 4th–6th grade; Adolescent version for students in 6th–12th grade) Supplemental survey material available for prevention planning

according to benefits from drinking, drinking styles, drinking consequences, and concerns about and recognition of a drinking problem. The primary scale factors include the following: (1) drinking to improve sociability; (2) drinking to improve mental functioning; (3) drinking to manage or change mood; (4) drinking to cope with marital problems; (5) gregarious versus solitary drinking; (6) obsessive-compulsive drinking or constantly thinking about drinking; (7) continuous, sustained drinking; (8) loss of behavior control when drinking; (9) social-role maladaptation; (10) perceptual withdrawal symptoms such as alcohol hallucinosis and delirium tremors; (11) somatic or physical withdrawal (e.g., shakes, hangovers, convulsions); (12) drinking provokes marital problems; (13) quantity of alcohol used; (14) post-drinking worry, fear, and guilt; (15) external support to stop drinking; (16) ready to quit; and (17) recognition of drinking problems.

The AUI primary scales often identify three general profiles of problem drinkers: low impairment problem drinkers, medium impairment problem drinkers, and high impairment problem drinkers (Rychtarik et al. 1998, 1999). Low impairment problem drinkers are likely to show a later onset of problem drinking and seek treatment as outpatients. They also are likely to be relatively successful in their social and vocational lives. The medium impairment problem drinkers are similar to the first type of drinker in that they show relatively good social adjustment. However, they are more likely to report a history of physical, emotional or sexual abuse, and depression. Finally, the high impairment drinkers

show the greatest social and vocational impairments, high levels of previous physical, emotional, or sexual abuse, highest levels of sustained drinking, and highest levels psychopathology (i.e., depression, anger, or sociopathy).

The *MacAndrew Alcoholism Scale/Revised* (MAC/MAC-R; MacAndrew 1965, 1989) is a subscale of the Minnesota Multiphasic Personality Inventory (MMPI), a standardized questionnaire developed by Hathaway and McKinley (1943) to assess psychopathology. The strength of this inventory is that it can be used to help rule out possible psychopathology when given in conjunction with the entire MMPI-2 scale—an important decision when investigating the cause of maladaptive behaviors such as drug misuse. However, the reliability of the scale is questionable for both females ($\alpha = 0.45$) and males ($\alpha = 0.56$). Some profiles characterize alcohol and/or drug abuse as a form of self-medication for depression (e.g., the 24/42 scale). The MAC/MAC-R consists of 49 items that differentiate between alcoholic patients and nonalcoholic psychiatric patients (Clopton 1978; Clopton et al. 1980; Svanum et al. 1982). The scale also has been found to help identify individuals who are at risk for developing alcohol-related problems (McCourt et al. 1971). One limitation of the scale is that it does not effectively differentiate alcohol abusers from other drug abusers (Burke and Marcus 1977). Additionally, female alcoholics consistently obtain higher scores than males with similar difficulties (Butcher and Owen 1978). Higher scores suggest potential drug abuse but are also suggestive of extraversion, assertiveness, risk-taking, and the possibility of having experienced blackouts and difficulty concentrating. Low scores are suggestive of introversion, conformity, and low self-confidence, as well as being contraindicative of drug abuse. The MAC was placed in the comprehensive assessment section assuming it is administered as part of the MMPI. Other subscales on the MMPI can help researchers or clinicians identify other psychological problems that are associated with or underlie drug misuse. Additional research with the scale led to speculations regarding the content and predictive validity for use with different populations and different cutoff scores (24 vs. 28) (Stein et al. 1999). Researchers have since developed additional, complimentary, forms for detecting alcoholism and drug abuse with the MMPI-2, the 39-item Addiction Potential Scale (APS) and the 13-item Addiction Acknowledgement Scale (AAS) (Weed et al. 1992).

The original *Chemical Dependency Assessment Profile* (*CDAP*) is a 232-item multiple-choice, true/false, and open-ended self-report questionnaire used to assess substance use, dependence problems and treatment needs among adolescents and adults (Davis et al. 1989). Further analysis of a modified version of the instrument—including only the multiple-choice and true/false items—revealed dimensions of dysfunction addressing quantity/frequency of use, physiological symptoms, situational stressors, antisocial behaviors, interpersonal problems, affective dysfunction, treatment attitudes, impact of use on life functioning, and expectancies (Harrell et al. 1991). Original findings indicate good internal consistency ($\alpha = 0.78–0.88$), but normative data was only based on a sample of 86 individuals. This assessment is unique in that it taps into neurobiologically based, cognitively based, and socially based drug use motivations and consequences for alcohol users, nonalcohol drug users and polydrug users.

The *Inventory of Drinking Situations* (IDS; Annis 1982) or the alternative *Inventory of Drug Use Situations* (Annis et al. 1992) assesses the contextual aspects of alcohol or other drug use and provides information about relapse situations. This inventory consists of either 42 or 100 items—with eight subscales—to evaluate drinking/drug use situations, including unpleasant emotions, physical discomfort, pleasant emotions, testing personal control, urges and temptations, conflict with others, social pressures, and pleasant times with others. While the measure does not directly measure alcohol misuse behavior, it provides a unique and comprehensive perspective into the ecological patterns of alcohol use behavior.

The *Substance Abuse Subtle Screening Inventory* (*SASSI*) consists of 81 items and ten scales (face valid alcohol, face valid other drugs, family-friends risk, attitudes, symptoms, obvious attributes, subtle attitudes, defensiveness, supplemental addiction measures, and correctional). It is not clear, however, that the different scales measure empirically distinct phenomena as indicated by high internal consistency across studies for direct but not indirect (subtle) measures (Feldstein and Miller 2007). The face valid content measures show the best convergence with interview-based measures on substance use impairment (Nishimura et al. 2001; Rogers et al. 1997). Studies on the specificity of SASSI suggest that it may produce a higher rate of false positives when used for diagnostic purposes (Feldstein and Miller 2007).

The *Drug Use Screening Inventory* (DUSI; Tarter 1990) is a self-report inventory which consists of 149 items (the revised version, the DUSI-R, consists of 159 items) and is used to quantify problems in ten areas, including alcohol or other drug use, behavior problems, health status, psychiatric disorders, social competence, family adjustment, school adjustment, peer relations, and leisure/recreational time. The scale demonstrated good internal consistency for males (KR20 = 0.76) and females (KR20 = 0.72) and has been shown to be a valid measure of drug misuse in normal and clinical populations (Tarter and Hegedus 1991). An adolescent version is also available (Tarter et al. 1992) that is adequate at discriminating between normal and clinical drug users based on the DSM-III criteria for a psychoactive substance use disorder (Kirisci et al. 1995).

11.6.1 Comprehensive Assessments for Adolescent Populations

The *Problem Oriented Screening Instrument for Teenagers* (*POSIT*) is a 139-item self-administered yes/no questionnaire that was developed by the National Institute on Drug Abuse as part of their Adolescent Assessment/Referral System (Rahdert 1991). The POSIT contains ten scales: substance use/abuse, physical health status, mental health status, peer relations, family relations, educational status, vocational status, social skills, leisure and recreation, and aggressive behavior/delinquency. This measure has good convergent validity, internal consistency, and test–retest reliability and it takes 20 min to complete. This measure takes into account all four etiologic domains presented in this text.

The *Personal Experience Inventory* (PEI; Winters and Henly 1989; Winters et al. 1993) is a multidimensional questionnaire used for detection of problem consequences and potential risk factors associated with diagnostic classification of substance use disorders in adolescent populations (Guthmann and Brenna 1990). This 276-item questionnaire helps to quantify level of involvement with a variety of drugs and the severity of problems in personal, family, and psychosocial domains. The scale also presents individuals with questions related to cognitive, social, and immediate environmental impacts of their drug use in addition to items concerning social reinforcement to maintain drug use. A variation of this inventory also exists for adults (Winters 1999).

The *American Drug and Alcohol Survey* (ADAS) is a 57-item questionnaire used by school systems to identify patterns of substance use and abuse (Oetting et al. 1999). Two versions of the scale are available depending on the age of the school children. The "Children's form" is used with students in 4th–6th grade and measures alcohol and drug prevalence (for five substances), lifetime/annual/past 30-day use, and information regarding peer, family and school variables associated with use. The "Adolescent's form" is used with students in 6th–12th grade and measures alcohol and drug prevalence (for 21 substances), lifetime/annual/past 30-day use, perceived availability, peer and family variables, drug use consequences, perceived harm, location of drug use, high-risk drug behavior, and future intent to use drugs. The scale has demonstrated good to excellent internal consistency (α = 0.72–0.97) and has been validated with minority

student populations (Ezell and Burrell 2011). Supplemental measures are also available to assist schools with conducting a needs assessment for preventive intervention and community readiness. A potential limitation of the scale is that the developer must perform scoring as no instructions are provided with testing materials; however, researchers can request a data file for further analysis.

11.6.2 Strengths and Weaknesses of Using Comprehensive Inventory Assessments

The primary advantage of employing an exhaustive inventory assessment is the comprehensive nature of the measures. This can be an advantageous alternative to interview assessments in clinical settings where the clinician to patient ratio is extremely low. These measures are also useful for researchers with limited time to develop survey batteries to collect detailed information in multiple domains of functioning. Furthermore, several of these assessments are also commonly used to assess other psychological illnesses that allows for the detection and treatment of those with comorbid conditions.

Unfortunately, many of these inventories consist of one hundred or more items that may lead respondents to disengage from the assessment and begin responding in random or incomprehensible patterns (Meade and Craig 2011). Moreover, these inventory assessments are also vulnerable to the self-report limitations described earlier, and the administrator may have to consider collecting corroborative information to verify the responses collected. Constructing a well-balanced and psychometrically valid comprehensive questionnaire is a tedious and time consuming task vulnerable to a number of biases that can weaken its validity as a diagnostic tool (Choi and Pak 2005). However, the data generated from comprehensive assessments can be extremely useful for detecting patterns of drug use problems over disparate combinations of drug use (i.e., polysubstance use). To date, there has been little work examining how patterns of drug use are related to specific patterns of drug use problems, but data generated from these measures can be very useful for this purpose. Understanding these patterns has the potential to increase the sensitivity and specificity of screening procedures in clinical settings that can ultimately influence diagnosis and prognosis for individual treatment regimens.

11.7 Biochemical Assessments of Drug Misuse

Biochemical screening methods can play an important role in assessment and treatment of adolescents and adults with substance use problems.[3] For initial drug use screening, the most commonly used tests are immunoassays (e.g., radioimmunoassay, enzyme immunoassay, and fluorescence polarization immunoassay). Immunoassays involve the measurement of labeled and unlabeled antigen (drug or metabolite) and antibody interactions (Goldberger and Jenkins 1999) collected from urine, blood, or hair samples. In drug testing, the antigen is a drug or metabolite and its corresponding labeled analog, and the antibody is a protein grown in an animal and directed toward a specific drug, metabolite, or group of similar compounds. More selective screening assays used for confirmation include gas chromatography/mass spectrometry (GC/MS), gas chromatography (GC), and high-performance liquid chromatography (HPLC). Chromatography consists of a variety of techniques used to separate mixtures of drugs, their metabolites, and other chemicals into individual components based on differences in relative affinity for a mobile phase and a stationary phase.

Positive test results for any substance are generally confirmed by a second test on the same urine sample, using a different analytic method. Alternative methods to urine analyses are hair, saliva, and blood analyses although these analyses are typically more expensive. These tests provide validation of the accuracy of

[3]See also Chap. 14 in this volume

self-reported substance use when properly conducted and when the results are properly interpreted to minimize errors (e.g., false-positive or false-negative test results). Previous work has found that self-report of drug use is concordant with immunoassay measurements in drug using populations (Denis et al. 2012), although others have found underreporting and over-reporting to be a function of drug of choice and population characteristics (Zanis et al. 1994; Chermack et al. 2000). Biochemical assessment of drug use may also be used among adolescents and emerging adults to initiate early treatment, to rule out other possible illness or potential health problems when individuals are brought into the emergency room, to facilitate fair play in sports and scholarship, and to provide legal reasons to prove one's innocence (see Sussman and Ames 2001).

The Breathalyzer measures the blood alcohol concentration (BAC) of an individual using expired carbon monoxide air samples. It is a commonly used tool by law enforcement to assess possible alcohol intoxication by drivers, and can be used as evidence in future legal proceedings. The Breathalyzer collects an exhaled air sample from the individual via a mouthpiece connected to a tube leading to a chamber in the device connected to a vile containing sulfuric acid, potassium dichromate, silver nitrate, and water. The sulfuric acid removes the alcohol from a gas to liquid state where it reacts with the potassium dichromate to a new chemical formulation. In professional grade models, fuel cells generate an electrical current to compare the chemical levels in the collection vial to a second vial of an unreacted mixture in the device. The higher the discordance in chemical status is evidence of a higher concentration of alcohol collected from the air. Research has evidenced that this technique correlates highly with concurrent blood samples ($r = 0.98$) (Peleg et al. 2010).

Expired carbon monoxide (CO) air samples can also be used to collect and immediately assess for tobacco use, with a relatively short half-life (3–5 h). Thiocyanates (SCN) are found in body fluids, partly as a result of detoxification of hydrogen cyanide in cigarette smoke (Luepker et al. 1981), and have a half-life of 10–14 days. However, SCN levels can be inflated by cyanogenic foods, such as cabbage, and can be influenced by factors that change intercellular fluid volume. The measurement of cotinine, a major metabolite of nicotine, is a more precise measure of nicotine intake and has a half-life of 30 h (2- to 4-day detection period). A positive test with CO paired with a negative cotinine test could indicate marijuana use.

11.7.1 Measuring Genetic Susceptibility

Over the past decade, researchers have taken an interest in the measurement of genetic vulnerability to drug use, abuse and addiction. While it is generally accepted that health behavior is a function of the interaction between genetic and environmental influences, researchers have begun to estimate the genetic influence for drug use behavior. Interestingly, the degree of genetic influence is specific to classes of drugs and due to variation in numerous genes in the human genome (see Kreek et al. 2005 for review). Research continues to identify genetic variations that contribute to drug use and abuse through genome-wide association studies (GWAS) (Bierut et al. 2010; Tobacco and Genetics Consortium 2010; Hall et al. 2013; Vrieze et al. 2013), and this work presents a unique opportunity for transdisciplinary efforts to preclude drug misuse as discussed below.

11.7.2 Strengths and Weaknesses of Using Biochemical Measures of Drug Misuse

Biochemical assessments provide accurate and objective information regarding the type and amount of a particular drug that is present in the body at the time of measurement. Urine immunoassays have been a popular biochemical method of assessing level of drug use as a cheap and effective method of corroborating self-report

information. However, this method is not without its limitations. Detection of drug metabolites is dependent on the sensitivity of the assay used. Drug concentrations are highest several hours after drug use and decrease to undetectable levels over time. The length of time a drug or its metabolites can be detected in urine is referred to as the *retention time*. Retention times differ according to (1) the type and amount of drug consumed, (2) whether use is occasional or chronic, (3) the method of drug use, (4) individual metabolic rates and excretion, (5) diet, (6) acidity of the urine, (7) fluid intake, and (8) the time of day (Moeller et al. 2008).

Generally, the length of time drugs stay in the body varies across drug types. For example, cocaine and some hallucinogens (e.g., LSD) are present in the body 12–48 h. Drugs that are present in the body 1–3 days include methadone, opiates (heroin, morphine, codeine), proposyphene (Darvon), methaqualone (Quaalude), barbiturates (e.g., Phenobarbital), and amphetamines (crystal, ice, crank, methamphetamines; 1–2 days). Phencyclidine (PCP) used occasionally remains present in the body for 1–8 days, whereas when chronic use is present, PCP remains in the body up to 30 days. Finally, cannabinoids (marijuana) used occasionally are present in the body for 1–7 days, whereas daily chronic use causes cannabinoids to remain present in the body for 1–6 weeks (Moeller et al. 2008).

Evidence has also suggested that false positive rates are more common with certain drugs, particularly benzodiazepine, than others—such as marijuana and crack cocaine use (Vincent et al. 2006). Prescription and nonprescription drug use may also contribute to false-positive rates and should be considered during the screening process (Brahm et al. 2010). In accordance with previous conclusions, it is recommended that additional "confirmatory" testing be performed in lieu of drawing definitive conclusions from urine-based assessments. Of course, these more comprehensive biological tests may not be available for rapid assessment that may limit the researcher in making a final decision about the participant's drug use status.

In addition to these limitations of popular biological drug screening methods, there has been little to no work that has attempted to determine how these measures correlate to the definition of drug misuse provided at the beginning of the chapter. This is a challenging task given the heterogeneity of biological responses to drugs between individuals. Whereas new drug users may achieve a high from a small amount of some drug, the experienced user may require a dose two to three times greater due to an accumulated biological tolerance. As such, while the biochemical assessment may indicate the experienced user has more drugs in their system, they may be experiencing fewer intrapersonal or interpersonal difficulties relative to a new user. Despite these limitations, the continued research of drug concentration levels is not without merit as it can be useful in predicting outcomes and improving treatment decisions for drug using patients in clinical and natural settings beyond self-report measures of drug use (Isbister 2010). For example, the installation of Breathalyzer enabled ignition interlock devices can assess the level of alcohol in a driver's system before allowing the car to operate. If judged to be beyond the legal limit, it can prevent the individual from driving under the influence of alcohol—a common and dangerous form of alcohol misuse.

11.8 Multimodal Assessment and Concordance of Self-Report, Interview, and Biochemical Assessments of Drug Misuse

Multimodal assessment is a technique where drug use data is collected concomitantly using self-report, biological and observational methods. The *portal survey technique* (Voas et al. 2006; Kelley-Baker et al. 2007) is a multimodal assessment procedure that is being primarily used in field research with high-risk alcohol, and other drug using participants. The rationale for developing this technique stemmed from limitations of

collecting self-report data emerging adult participants (Kelley-Baker et al. 2007). The obvious limitation is that self-report data tends to underestimate individual and population level drug use as survey responses are collected anonymously to protect the confidentiality of participants. This measurement strategy is useful in field research —for example, emerging adult populations are difficult to track and generally neglected in research with the exception of samples that come from college attending or stable, employed individuals. Moreover, the portal survey method is especially useful in environments known to encourage alcohol or other drug use—defined as high-risk settings—where the where access to participants is defined and limited by the setting in question (i.e., bar, sporting event, night club, etc.). Multimodal assessment techniques are gaining popularity in field research for the reasons outlined above, but the research community would benefit from understanding the potential weaknesses of this approach.

Perhaps the issue of greatest concern surrounds concordance rates of measures (self-report vs. biological vs. interview). When discordant information is collected, the researcher must decide what information can be considered valid and used for statistical analysis. The study of concordance rates emerged as scientists attempted to develop biological measures to detect drug use to use as a corroborative form of evidence when collecting self-reported drug use. Despite increased sophistication in biological assessments, concordance rates are seldom, if ever, perfect. This is especially concerning in "high risk" environments where multimodal assessment would be used. For instance, Johnson et al. (2009) collected self-report and saliva samples from young adults attending an electronic dance music event (i.e., a rave) where drug use is prevalent. They measured cocaine, marijuana, and amphetamine use—due to their prevalence in club settings—and found that only 41% of participants with drugs present in their system reported drug use. Encountering discordant information between biological and self-report measures of drug use can be attributed to several factors working independently or in combination. First, the honesty of the participant may be a function of individual drug use pattern. For example, Ledgerwood et al. (2008) found that biological assessment, via a hair sample, was more accurate than self-report at detecting past 90-day cocaine use but was not more advantageous than self-report when measuring marijuana, opiates, or methamphetamine use. Of course, this limitation is influenced by larger sociopolitical factors. In particular, it is likely that individuals are less inclined to self-report illicit "hard" drug use given the legal ramifications and social stigma associated with use. Research investigating more commonly used and socially acceptable drugs (e.g., alcohol, tobacco, marijuana) has yielded stronger concordance rates between self-report and biological measurements (Nichols et al. 2014). Second, some biological assessments, such as hair samples, are not accurate at detecting polydrug use unless a large sample is taken (Ledgerwood et al. 2008). Third, biological assessments are not ideal in households, or other environments, where non-drug using individuals are exposed to airborne byproducts of drug use as contaminants can lead to inflated false-positive rates (Delaney-Black et al. 2010). Although this is not true for all biological assessment tools this should be taken into consideration by researchers prior to analysis.

Similar concerns exist when comparing self-report measures to face-to-face interview collection strategies. Stone and Latimer (2005) investigated the concordance rates of these strategies in a sample of adolescents and found strong correlations in the reporting of alcohol and marijuana use with average correlations of 0.72 and 0.81, respectively. However, they noted that participants reported higher frequency of use when data was collected in an interview format. One may conjecture that participants are more likely to respond honestly when they believe the

method demonstrates a more direct "pipeline" to the truth (e.g., interview or biochemical) compared to self-report questionnaires (Sussman and Ames 2008).

11.9 Practical and Ethical Issues to Consider When Measuring and Assessing Drug Misuse from Secondary Data Sources

The collection and analysis of drug use data poses several ethical and practical challenges related to the source of the data, the characteristics of the patients/participants and in choosing appropriate measures to achieve research goals. While U.S. government-funded national datasets are available to researchers at no charge, accessing and utilizing data from clinical settings poses several challenges. Trust issues between community-based organizations and research institutions are not a new phenomenon. In fact, many have described trust issues and provided recommendations for quelling possible problems that arise in research-community collaborations (Israel et al. 2005; Christopher et al. 2008) of which a few are discussed below. Establishing rapport with community-based treatment centers, not-for-profit or for-profit medical centers are a challenging task. More often than not, a gap in objectives exists. Treatment facilities are interested in delivering timely and effective care to those with a substance use problem whereas a researcher is more often interested in testing theory-driven or empirically supported interventions, evaluating the prevalence of comorbid conditions in clinical populations or developing statistical models evaluating mediation and moderation effects for treatment success or failure. Establishing common goals is recommended as a good initial step to building relationships and gaining access to data from clinical populations. It is also important for researchers to acknowledge the expertise of their collaborators to encourage them to participate in the development and dissemination of research procedures. Identifying individual strengths or recognizing personal or institutional histories (i.e., accomplishments related to public health) can also motivate community partners to collaborate. Finally, being present—and engaged—with collaborating institutions serves to enhance rapport with officials and prospective research participants.

Recent changes in the U.S. drug policy have improved the communication channel between clinicians and researchers. Recent changes via the Affordable Care Act (ACA) present a unique opportunity for researchers and clinicians to engage in shared goals to understand and improve delivery of treatment in publically funded substance abuse treatment centers (Buck 2011). It is likely that the demand for the dissemination and evaluation of evidence-based treatments will grow in the coming years. This poses a unique opportunity for collaboration as clinicians are in need of researchers with strong methodological and statistical skills to aid in this endeavor, while researchers remain in need of an existing infrastructure to collect and analyze alcohol and drug misuse data from clinical populations. These policy changes are likely to foster new, and improve existing, clinician and researcher relationships through the development of aims that align the goals of clinical and research work detecting preventing and treating drug misuse, abuse, and addiction.

While endeavors such as those outlined above are likely to give researchers broader access to clinical data, it behooves the research community to remain cognizant of federal regulations pertaining to the use of data collected without consent or intention of being analyzed for research (i.e., non-patient care) purposes. It is recommended that researchers advise policy makers regarding appropriate federal and state laws governing the protection of medical and behavioral health records. Federal laws, most notably the Health Insurance Portability and Accountability Act (HIPAA) of 1996, provide a framework for protecting patient privacy and confidentiality. While a full discussion of HIPAA

regulations is beyond the scope of this chapter, a few points relevant to hypothesis testing and general data analysis are necessitated.

Most importantly, clinical data used for research purposes should be deidentified. Deidentification of data involves the removal of any personal health information (PHI) that could be used to identify an individual in a dataset. Eighteen specific identifiers are listed under the HIPAA law that must be removed to be considered deidentified including: (1) name, (2) geographical information smaller than State (e.g., street address, zip code, city, county), (3) dates (e.g., birth, admission, death, etc.), (4) phone numbers, (5) fax numbers, (6) e-mail address, (7) social security address, (8) medical record numbers, (9) health plan beneficiary numbers, (10) account numbers, (11) certificate/license numbers, (12) vehicle identifiers (e.g., license plate number) and serial numbers, (13) device identifiers and serial numbers, (14) URLs, (15) IP address numbers, (16) biometric identifiers (e.g., thumb print), (17) full face photographic images, and (18) any other unique identifying number, characteristic or code. Additional data may be needed to conduct spatial analyses or to explore topics such as length of time from diagnosis to treatment or time from treatment to relapse. In such instances, a limited dataset including information such as zip codes (residential or treatment facility) and dates (diagnosis, treatment, etc.) can be used so long as (1) direct identifiers (name, address, etc.) are removed and (2) a data use agreement is established to prohibit activities such as attempting to identify or contact individuals. Regardless, researchers should prescribe to the "minimum necessary standard" by limiting the disclosure of PHI to the minimum amount necessary to answer the research questions. It is of critical importance to recognize that, in the U.S., individual states may have additional privacy laws that could apply to research endeavors using clinical data. As a rule of thumb, researchers should submit to the more restrictive law (state or federal) when collecting and analyzing drug use and other health-related data.

In addition to meeting the standards of national and state laws on secondary data analysis, it is advisable for researchers to enter into a memorandum of understanding (MOU) with any community-based, private or publically funded treatment center prior to commencing collaborative work involving the exchange of data. The MOU is a formal agreement between two parties intended to establish an official partnership. While an MOU is not a legally binding document, it serves to outline the objectives of the collaboration, responsibilities of the involved parties, any financial agreements, and issues related to data ownership, use and dissemination of research results. Encouraging the use of an MOU can help to promote confidence in the researcher's sincerity to enter a mutually beneficial relationship.

In some cases, researchers are advised to consider the use of certificates of confidentiality (COC) when collecting, or accessing, subject identifiable data on illicit drug use or related illegal activities. The COC is designed to "prevent consequential harms associated with compulsory legal disclosure of identifiable research data" (Currie 2005, p. 7). Distributed by the NIH, the COC can enhance the privacy of self-report data by giving researchers the ability to not disclose information if such information can damage the financial standing, legal standing, insurability, or reputation of the participant. Applying for a COC should be considered when working with underage smokers and drinkers, when collecting data on illicit drug use, or on risky health behaviors associated with drug use (i.e., driving under the influence, risky sex, intoxicated physical fights, etc.). Researchers can apply for a COC regardless of the funding mechanism for the research project provided the research gains approval from a human subjects review committee (IRB). Of course, a COC is limited in that certain information can be disclosed to external entities if voluntarily reported, including: child abuse, threats to harm self or others, or contracting a reportable communicable disease (e.g., HIV).

Analyzing existing datasets, such as publically available data or data collected in medical settings, is a worthwhile alternative to conducting an original research project; however, it is not advisable without first understanding the strengths and limitations of performing a secondary data analysis. First, the researcher should remain vigilant to issues related to study design. Statistical conclusion validity is threatened when researchers neglect issues of design, temporal ordering of events and sampling procedures (i.e., representativeness) especially in cases where causal associations or generalizations to populations are desirable. On a related note, researchers should consider the nature of the measures employed in the collection of primary data to discern whether the data qualifies as substance *misuse* data and ensure that any proxy measures adequately reflect the psychological or behavioral constructs of interest. Shamblen and Dwivedi (2010), for example, demonstrated this in their finding that responses to items of excessive alcohol use (i.e., binge drinking) were more consistent over time compared to normally distributed indicators of drinking behavior which lead to unexpected, and counterintuitive, relationships with other variables of interest. Second, the dataset must be scrutinized for completeness, accuracy, consistency of measurement, timeliness of assessments (for longitudinal data) and other data-related issues in order to preserve adequate internal validity to draw cause-and-effect relationships. While many national datasets maintain a detailed record of study design and issues related to statistical procedures for variable calculations and missing data, data collected from small organizations may require "cleaning" and recoding to ensure its consistency and completeness. Finally, and perhaps most importantly, the researcher must remain mindful of what research questions can, and cannot be addressed with secondary data. Vartanian (2010), for instance, posits that using secondary data undermines the research process since the research questions can only be expressed in terms of the existing data—a process known as "driving the question." In sum, he recommends secondary data only be used for exploratory work or hypothesis generation, but not for formal hypothesis testing.

Despite these limitations, there are several advantages to using secondary datasets that should not go without mention. First, it is a relatively low-cost endeavor, as no funds are needed for data collection and participant retention. Second, issues surrounding inadequate sample size or population representativeness typically do not plague large nationwide studies. Finally, several secondary datasets have been derived from information collected in clinical settings. Data collected in clinical settings is less influenced by biases common in research settings (e.g., recall bias, nonresponse bias, experimenter bias, etc.), although some bias is inherent in all data collection and should be addressed as needed (see Boslaugh 2007).

11.10 Current and Future Transdisciplinary Assessment Efforts

Appraising an individual as engaging in drug misuse depends on the combination of various biological, intrapersonal, interpersonal, and social factors. The operationalization of drug misuse that has guided the discussion in this chapter may undermine individual differences (e.g., biological susceptibility, social circumstances, personality characteristics, etc.) in the consumption—and escalation—of drug use, the intrapersonal and interpersonal consequences experienced, and the broader social impact of individual drug use.

At present, the majority of work in the area of drug misuse assessment has reflected a multidisciplinary approach (whereby researchers from different disciplines work independently from their own conceptual framework to address a common problem) as opposed to a transdisciplinary approach (whereby researchers from different disciplines work jointly using a shared conceptual framework to address a common problem). Of course, these efforts are not without importance as it has generated findings that have aided in the assessment of drug misuse. For

example, the development of an estimated Blood Alcohol Concentration (eBAC) equation (Matthews and Miller 1979) has been found to correlate with actual intoxication (Hustad and Carey 2005), and allows researchers to produce a reliable proxy measure of the biological level of alcohol use without the need to collect a biological specimen or rely on a participant's subjective estimate.

The focus of most transdisciplinary efforts to date has been the melding of biological and psychosocial measures of drug misuse with an emphasis on genetic influences (Jang et al. 1995, 2001; Lynskey et al. 2012). Such studies have redefined our understanding of susceptibility to drug use and misuse as being, at least in part, heritable. Some have even surmised that heritability can contribute up to 50% of our susceptibility to the effects of drugs (Uhl et al. 2008). An ongoing challenge for researchers and clinicians remains in understanding the heritability of drug use and misuse in order to improve upon screening and treatment protocols. Subdisciplines in the field of bioinformatics have made great strides in identifying some genetic components related to drug misuse potential and how it may effect treatment options (Kreek et al. 2005; Mroziewicz and Tyndale 2010), but it remains a task for researchers in other health fields to translate this information in order to improve current screening, treatment, and recovery processes for those misusing drugs. Insofar as examining the genetic and environment interactions, future work would benefit from continuing to refine measures of environment when examining multiple levels of influence on drug use and misuse (Vrieze et al. 2012).

Provided the rather nascent state of transdisciplinary work in the field of drug misuse assessment, it is recommended that future transdisciplinary efforts in the study of drug misuse aim to coalesce biological, psychological, and social/environmental measures to (1) enhance our understanding of drug misuse and aid in the development of an operational definition that can be applied across individuals, social groups, and cultures, (2) improve our understanding of individual biological differences as it relates to the intrapersonal and interpersonal consequences of drug misuse in the context of the social and built environment, (3) develop improved screening techniques to identify and classify drug misuse, and (4) develop novel measures that can aid researchers and clinicians to provide improved treatment options for those with a diagnosable drug use disorder.

11.11 Summary, Future Directions, and Conclusions

Research on the assessment of drug misuse is essential developing later applications, including evaluations for an individual's treatment needs and ruling out other potential reasons for behavioral changes. The present chapter provided information on commonly used techniques of drug misuse assessment, the strengths and weaknesses of different approaches, multimodal assessment, issues related to field research, collecting data from clinical settings, the strengths and limitations of secondary data analysis and concluded with thoughts on transdisciplinary efforts for the field. To date, many interview and self-report assessment methods have been developed and implemented to help quantify drug use behavior and the severity of consequences associated with drug use. While this chapter does not provide an exhaustive list of assessments, it does highlight several well-known techniques for evaluating drug misuse that can aid in the detection of drug misuse future substance use disorders.

Implicit to the diverse methods employed to assess drug misuse is the underlying issue of defining and operationalizing drug misuse. The inferred definition of drug misuse among health and medical professionals has been largely driven by (1) the idea of misuse as it pertains to prescription drugs and (2) advances in neurobiological studies of addiction in animals and humans. Prescription drug *misuse* is commonly described as the use of a prescription medication without a written prescription from a physician or the use of a prescription medication in a way other than prescribed. The National Institute on

Drug Abuse (NIDA) and other health organizations have provided a similar definition for prescription drug misuse (or abuse), and the interpretation that misuse and abuse are synonymous has driven the theoretical understanding of alcohol and illicit drug misuse. Conversely, recent progress in understanding the neurobiological effects of drug use has redefined our understanding of the progression toward addiction. In fact, addiction has come to be defined as a progressive and relapsing brain disease (NIDA 2010) that has effects on, and is affected by, several regions of the brain that has both genetic and environmental determinants (Koob and Le Moal 2006; Kalivas and Volkow 2005).

Social and behavioral research has been influenced by the ideas such as those put forth in the medical field, leading drug misuse to commonly be described as either a distinct stage in the progression from experimentation to drug abuse/addiction/dependency or a singular characteristic in determining the severity of a drug use disorder. These conceptualizations have focused on earlier work around problematic alcohol use that was defined, and measured, by the proximal "short-term" effects of excessive alcohol use. Describing drug misuse as a transitional stage between occasional or regular use and abuse or addiction warrants merit, but is better defined by ecological limitations than its theoretical strengths. For example, an individual may be experiencing a plethora of intrapersonal and interpersonal consequences from continued drug use, but may not be exposed to a treatment setting due to a lack of health insurance or other pertinent factors (e.g., fear of stigma). Without a formal diagnosis, they would not be identified as a "drug abuser" but exhibit the characteristics of drug misuse. Theoretically, the identification of drug misuse can progress into screening for a formal drug use disorder diagnosis or allow the individual to consider intervention strategies to reduce their drug use before the culmination of proximal drug use consequences warrants such a diagnosis. As a "precursor" of drug abuse, drug misuse is an assessment of the ongoing consequences of drug use in addition to other important factors including, but not limited to, frequency and duration of drug use and level of chemical dependency. While no single unified definition of drug misuse is currently recognized across scientific and clinical disciplines, advances in transdisciplinary work can aid in the development of an operationalized definition that incorporates biological, social, and psychological components. Such work is necessary to our understanding of addiction, identifying resources to prevent it, and developing tailored interventions to treat the disease and promote resilience.

References

Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the alcohol use disorders identification test (AUDIT). *Alcoholism, Clinical and Experimental Research, 21*(4), 613–619.

American Psychiatric Association (Ed.). (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR®*. Washington, DC: American Psychiatric Publications.

American Psychiatric Association. (2006). *American psychiatric association practice guidelines for the treatment of psychiatric disorders: Compendium*. Arlington, VA: American Psychiatric Press.

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-V*. Washington, DC: American Psychiatric Publications.

Annis, H. M. (1982). *Inventory of drinking situations (IDS-100)*. Toronto, CAN: Addiction Research Foundation of Ontario.

Annis, H. N., Martin, G., & Graham, J. M. (1992). *Inventory of drug taking situations (alcohol). IDTSA. User's guide*. CAN: Addiction Research Foundation.

Babor, T. F., Hofmann, M., Del Boca, F. K., Hesselbrock, V., Meyer, R. E., Dolinsky, Z. S., et al. (1992). Types of alcoholics, I: Evidence for an empirically derived typology based on indicators of vulnerability and severity. *Archives of General Psychiatry, 49*(8), 599–608.

Bierut, L. J., Agrawal, A., Bucholz, K. K., Doheny, K. F., Laurie, C., Pugh, E., et al. (2010). A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences, 107*(11), 5082–5087.

Blanchard, K. A., Morgenstern, J., Morgan, T. J., Lobouvie, E. W., & Bux, D. A. (2003). Assessing

consequences of substance use: Psychometric properties of the inventory of drug use consequences. *Psychology of Addictive Behaviors, 17*(4), 328–331.

Boslaugh, S. (2007). *Secondary data sources for public health: A practical guide*. New York, NY: Cambridge University Press.

Brahm, N. C., Yeager, L. L., Fox, M. D., Farmer, K. C., & Palmer, T. A. (2010). Commonly prescribed medications and potential false-positive urine drug screens. *American Journal of Health-System Pharmacy, 67*(16), 1344–1350.

Buck, J. A. (2011). The looming expansion and transformation of public substance abuse treatment under the Affordable Care Act. *Health Affairs, 30*(8), 1402–1410.

Burke, H. R., & Marcus, R. (1977). MacAndrew MMPI alcoholism scale: Alcoholism and drug addictiveness. *The Journal of Psychology, 96*(1), 141–148.

Butcher, J. N., & Owen, P. L. (1978). Objective personality inventories: Recent research and some contemporary issues. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook* (pp. 475–545). New York, NY: Springer.

Buu, A., Li, R., Walton, M. A., Yang, H., Zimmerman, M. A., & Cunningham, R. M. (2014). Changes in substance use-related health risk behaviors on the timeline follow-back interview as a function of length of recall period. *Substance Use and Misuse, 49*(10), 1259–1269.

Chavez, P. R., Nelson, D. E., Naimi, T. S., & Brewer, R. D. (2011). Impact of a new gender-specific definition for binge drinking on prevalence estimates for women. *American Journal of Preventive Medicine, 40*(4), 468–471.

Chermack, S. T., Roll, J., Reilly, M., Davis, L., Kilaru, U., & Grabowski, J. (2000). Comparison of patient self-reports and urinalysis results obtained under naturalistic methadone treatment conditions. *Drug and Alcohol Dependence, 59*(1), 43–49.

Choi, B. C. K., & Pak, A. W. P. (2005). A catalog of biases in questionnaires. *Preventing Chronic Disease, 2*(1), 1–13.

Christopher, S., Watts, V., McCormick, A. K. H. G., & Young, S. (2008). Building and maintaining trust in a community-based participatory research partnership. *American Journal of Public Health, 98*(8), 1398–1408.

Clopton, J. R. (1978). Alcoholism and the MMPI; A review. *Journal of Studies on Alcohol and Drugs, 39* (9), 1540–1558.

Clopton, J. R., Weiner, R. H., & Davis, H. G. (1980). Use of the MMPI in identification of alcoholic psychiatric patients. *Journal of Consulting and Clinical Psychology, 48*(3), 416–417.

Connors, G. J., & Volk, J. (2003). Self-report screening for alcohol problems among adults. In J. P. Allen & V. B. Wilson (Eds.), *Assessing alcohol problems a guide for clinicians and researchers*. NIAAA publication no. 03-3745. Washington, DC: U.S. Department of Health and Human Services.

Currie, P. M. (2005). Balancing privacy protections with efficient research: Institutional review boards and the use of certificates of confidentiality. *IRB: Ethics and Human Research, 27,* 7–12.

Davis, E., Harrell, T. H., & Honaker, L. M. (1989). Content domains of dysfunction in alcohol and polydrug abusers: The chemical dependency assessment profile. Paper presented at the meeting of the Association for the Advancement of Behavior Therapy, Washington, DC.

Delaney-Black, V., Chiodo, L. M., Hannigan, J. H., Greenwald, M. K., Janisse, J., Patterson, G., et al. (2010). Just say "I don't": Lack of concordance between teen report and biological measures of drug use. *Pediatrics, 126*(5), 887–893.

Denis, C., Fatséas, M., Beltran, V., Bonnet, C., Picard, S., Combourieu, I., et al. (2012). Validity of the self-reported drug use section of the addiction severity index and associated factors used under naturalistic conditions. *Substance Use and Misuse, 47*(4), 356–363.

Dennis, M. L., Funk, R., Godley, S. H., Godley, M. D., & Waldron, H. (2004). Cross-validation of the alcohol and cannabis use measures in the global appraisal of individual needs (GAIN) and timeline followback (TLFB; Form 90) among adolescents in substance abuse treatment. *Addiction, 99*(s2), 120–128.

Earleywine, M., LaBrie, J. W., & Pedersen, E. R. (2008). A brief Rutgers alcohol problem index with less potential for bias. *Addictive Behaviors, 33*(9), 1249–1253.

Ewing, J. A. (1984). Detecting alcoholism: The CAGE questionnaire. *Journal of the American Medical Association, 252*(14), 1905–1907.

Ezell, R. J., & Burrell, B. (2011). Predicting substance abuse of high school through purpose in life and religiosity. *International Journal of Humanities and Social Science, 1,* 111–126.

Feldstein, S. W., & Miller, W. R. (2007). Does subtle screening for substance abuse work? A review of the substance abuse subtle screening inventory (SASSI). *Addiction, 102*(1), 41–50.

Forcehimes, A. A., Tonigan, J. S., Miller, W. R., Kenna, G. A., & Baer, J. S. (2007). Psychometrics of the drinker inventory of consequences (DrInC). *Addictive Behaviors, 32*(8), 1699–1704.

Friedman, A. S., & Utada, A. (1989). A method for diagnosing and planning the treatment of adolescent drug abusers (the Adolescent drug abuse diagnosis [ADAD] instrument). *Journal of Drug Education, 19* (4), 285–312.

Gibbs, L. E. (1983). Validity and reliability of the Michigan Alcoholism Screening Test: A review. *Drug and Alcohol Dependence, 12*(3), 279–285.

Ginzler, J. A., Garrett, S. B., Baer, J. S., & Peterson, P. L. (2007). Measurement of negative consequences of substance use in street youth: An expanded use of the Rutgers alcohol problem index. *Addictive Behaviors, 32*(7), 1519–1525.

Goldberger, B. A., & Jenkins, A. J. (1999). Drug toxicology. In P. J. Ott, R. E. Tarter, & R. T. Ammerman (Eds.), *Sourcebook on substance abuse: Etiology,*

epidemiology, assessment, and treatment. Boston, MA: Allyn and Bacon.

Grigsby, T. J., Forster, M., Baezconde-Garbanati, L., Soto, D. W., & Unger, J. B. (2014). Do adolescent drug use consequences predict externalizing and internalizing problems in emerging adulthood as traditional drug measures do? *Addictive Behaviors, 39*(3), 644–651.

Grønkjær, M., Curtis, T., De Crespigny, C., & Delmar, C. (2011). Acceptance and expectance: Cultural norms for alcohol use in Denmark. *International Journal of Qualitative Studies on Health and Well-Being, 6*(4), 8461.

Guthmann, D. R., & Brenna, D. C. (1990). The personal experience inventory: An assessment of the instrument's validity among a delinquent population in Washington state. *Journal of Child & Adolescent Substance Abuse, 1*(2), 15–24.

Hall, F. S., Drgonova, J., Jain, S., & Uhl, G. R. (2013). Implications of genome wide association studies for addiction: Are our *a priori* assumptions all wrong? *Pharmacology & Therapeutics, 140*(3), 267–279.

Harrell, T. H., Honaker, L. M., & Davis, E. (1991). Cognitive and behavioral dimensions of dysfunction in alcohol and polydrug abusers. *Journal of Substance Abuse, 3*(4), 415–426.

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., et al. (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry, 170*(8), 834–851.

Hathaway, S. R., & McKinley, J. C. (1943). *The minnesota multiphasic personality inventory* (2nd Printing). New York, NY: Psychological Corporation.

Hjorthøj, C. R., Hjorthøj, A. R., & Nordentoft, M. (2012). Validity of timeline follow-back for self-reported use of cannabis and other illicit substances—Systematic review and meta-analysis. *Addictive Behaviors, 37*(3), 225–233.

Horn, J. L., Wanberg, H. W., & Foster, F. M. (1986). *The alcohol use inventory (AUI)*. Minneapolis, MN: National Computer Systems.

Hurlbut, S. C., & Sher, K. J. (1992). Assessing alcohol problems in college students. *Journal of American College Health, 41*(2), 49–58.

Hustad, J. T. P., & Carey, K. B. (2005). Using calculations to estimate blood alcohol concentrations for naturally occurring drinking episodes: A validity study. *Journal of Studies on Alcohol, 66*(1), 130–138.

Isbister, G. K. (2010). How do we use drug concentration data to improve the treatment of overdose patients? *Therapeutic Drug Monitoring, 32*(3), 300–304.

Israel, B. A., Parker, E. A., Rowe, Z., Salvatore, A., Minkler, M., López, J., et al. (2005). Community-based participatory research: Lessons learned from the centers for children's environmental health and disease prevention research. *Environmental Health Perspectives, 113*(10), 1463–1471.

Jang, K. L., Livesley, W., & Vernon, P. A. (1995). Alcohol and drug problems: A multivariate behavioural genetic analysis of co-morbidity. *Addiction, 90*(9), 1213–1221.

Jang, K. L., Vernon, P. A., Livesley, W. J., Stein, M. B., & Wolf, H. (2001). Intra-and extra-familial influences on alcohol and drug misuse: A twin study of gene–environment correlation. *Addiction, 96*(9), 1307–1318.

Johnson, M. B., Voas, R. A., Miller, B. A., & Holder, H. D. (2009). Predicting drug use at electronic music dance events: Self-reports and biological measurement. *Evaluation Review, 33*(3), 211–225.

Jones, K. D. (2010). The unstructured clinical interview. *Journal of Counseling & Development, 88*(2), 220–226.

Kahler, C. W., Strong, D. R., & Read, J. P. (2005). Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: The brief young adult alcohol consequences questionnaire. *Alcoholism, Clinical and Experimental Research, 29*(7), 1180–1189.

Kalivas, P. W., & Volkow, N. D. (2005). The neural basis of addiction: A pathology of motivation and choice. *American Journal of Psychiatry, 162*(8), 1403–1413.

Kelley-Baker, T., Voas, R. B., Johnson, M. B., Furr-Holden, C. D. M., & Compton, C. (2007). Multimethod measurement of high-risk drinking locations extending the portal survey method with follow-up telephone interviews. *Evaluation Review, 31*(5), 490–507.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005a). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*(6), 593–602.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005b). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*(6), 617–627.

Kirisci, L., Mezzich, A., & Tarter, R. (1995). Norms and sensitivity of the adolescent version of the drug use screening inventory. *Addictive Behaviors, 20*(2), 149–157.

Knight, J. R., Sherritt, L., Shrier, L. A., Harris, S. K., & Chang, G. (2002). Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. *Archives of Pediatrics and Adolescent Medicine, 156*(6), 607–614.

Knight, J. R., Shrier, L. A., Bravender, T. D., Farrell, M., Vander Bilt, J., & Shaffer, H. J. (1999). A new brief screen for adolescent substance abuse. *Archives of Pediatrics and Adolescent Medicine, 153*(6), 591–596.

Koob, G. F., & Le Moal, M. (2006). *Neurobiology of addiction*. London, UK: British Library Catalogue.

Krasnova, I. N., & Cadet, J. L. (2009). Methamphetamine toxicity and messengers of death. *Brain Research Reviews, 60*(2), 379–407.

Kreek, M. J., Nielsen, D. A., Butelman, E. R., & LaForge, K. S. (2005). Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nature Neuroscience, 8*(11), 1450–1457.

Ledgerwood, D. M., Goldberger, B. A., Risk, N. K., Lewis, C. E., & Kato Price, R. (2008). Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addictive Behaviors, 33*(9), 1131–1139.

Leonhard, C., Mulvey, K., Gastfriend, D. R., & Shwartz, M. (2000). The addiction severity index: A field study of internal consistency and validity. *Journal of Substance Abuse Treatment, 18*(2), 129–135.

Littrell, J. (1991). *Understanding and treating alcoholism: Biological, psychological, and social aspects of alcohol consumption and abuse* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Luepker, R. V., Pechacek, T. F., Murray, D. M., Johnson, C. A., Hund, F. R. E. D., & Jacobs, D. R. (1981). Saliva thiocyanate: A chemical indicator of cigarette smoking in adolescents. *American Journal of Public Health, 71*(12), 1320–1324.

Lynskey, M. T., Agrawal, A., Henders, A., Nelson, E. C., Madden, P. A., & Martin, N. G. (2012). An Australian twin study of cannabis and other illicit drug use and misuse, and other psychopathology. *Twin Research and Human Genetics, 1*(1), 1–11.

MacAndrew, C. (1965). The differentiation of male alcoholic outpatients from nonalcoholic psychiatric outpatients by means of the MMPI. *Quarterly Journal of Studies on Alcohol, 28*(2), 238–246.

MacAndrew, C. (1989). Factors associated with the problem-engendering use of substances by young men. *Journal of Studies on Alcohol and Drugs, 50*(6), 552–556.

Mäkelä, K. (2004). Studies of the reliability and validity of the addiction severity index. *Addiction, 99*(4), 398–410.

Marlow, D., & Crowne, D. P. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology, 25*(2), 109–115.

Matthews, D. B., & Miller, W. R. (1979). Estimating blood alcohol concentration: Two computer programs and their applications in therapy and research. *Addictive Behaviors, 4*(1), 55–60.

McCourt, W. F., Williams, A. F., & Schneider, L. (1971). Incidence of alcoholism in a state mental hospital population. *Quarterly Journal of Studies on Alcohol, 32*(4), 1085–1088.

McHugh, R. K., Sugarman, D. E., Kaufman, J. S., Park, S., Weiss, R. D., & Greenfield, S. F. (2014). Readability of self-report alcohol misuse measures. *Journal of Studies on Alcohol and Drugs, 75,* 328–334.

McLellan, A. T., Luborsky, L., Cacciola, J., Griffith, J., Evans, F., Barr, H. L., et al. (1985). New data from the addiction severity index reliability and validity in three centers. *The Journal of Nervous and Mental Disease, 173*(7), 412–423.

McLellan, A. T., Luborsky, L., Woody, G. E., & O'Brien, C. P. (1980). An improved diagnostic evaluation instrument for substance abuse patients: The addiction severity index. *The Journal of Nervous and Mental Disease, 168*(1), 26–33.

Meade, A. W., & Craig, S. B. (2011). Identifying careless responses in survey data. Paper presented at the 26th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Meyers, K., Hagan, T. A., McDermott, P., Webb, A., Randall, M., & Frantz, J. (2006). Factor structure of the comprehensive adolescent severity inventory (CASI): Results of reliability, validity, and generalizability analyses. *The American Journal of Drug and Alcohol Abuse, 32*(3), 287–310.

Meyers, K., Thomas McLellan, A., Jaeger, J. L., & Pettinati, H. M. (1995). The development of the comprehensive addiction severity index for adolescents (CASI-A): An interview for assessing multiple problems of adolescents. *Journal of Substance Abuse Treatment, 12*(3), 181–193.

Miele, G. M., Carpenter, K. M., Smith Cockerham, M., Dietz Trautman, K., Blaine, J., & Hasin, D. S. (2000a). Concurrent and predictive validity of the substance dependence severity scale (SDSS). *Drug and Alcohol Dependence, 59*(1), 77–88.

Miele, G. M., Carpenter, K. M., Smith Cockerham, M., Trautman, K. D., Blaine, J., & Hasin, D. S. (2000b). Substance dependence severity scale (SDSS): Reliability and validity of a clinician-administered interview for DSM-IV substance use disorders. *Drug and Alcohol Dependence, 59*(1), 63–75.

Miller, W. R., & Del Boca, F. K. (1994). Measurement of drinking behavior using the form 90 family of instruments. *Journal of Studies on Alcohol and Drugs, Supplement, 12,* 112–118.

Miller, W. R., Leckman, A. L., Delaney, H. D., & Tinkcom, M. (1992). Long-term follow-up of behavioral self-control training. *Journal of Studies on Alcohol and Drugs, 53*(3), 249–261.

Miller, W. R., & Marlatt, G. A. (1984). *Brief drinker profile*. Odessa, FL: Psychological Assessment Resources.

Miller, W. R., & Marlatt, G. A. (1987). *Manual supplement for the brief drinker profile, follow-up drinker profile, collateral interview form*. Odessa, FL: Psychological Assessment Resources.

Miller, W. R., Tonigan, J. S., & Longabaugh, R. (1995). The drinker inventory of consequences (DrInC). In *Project MATCH monograph series* (Vol. 4).

Moeller, K. E., Lee, K. C., & Kissack, J. C. (2008). Urine drug screening: Practical guide for clinicians. *Mayo Clinic Proceedings, 83*(1), 66–76.

Mroziewicz, M., & Tyndale, R. F. (2010). Pharmacogenetics: A tool for identifying genetic factors in drug dependence and response to treatment. *Addiction Science & Clinical Practice, 5*(2), 17–29.

National Institute of Drug Abuse. (2010). *Understanding drug abuse and drug addiction*. Rockville, MD: NIDA. Available at https://www.drugabuse.gov/sites/default/files/drugfacts_understanding_addiction_final_0.pdf

National Institute on Alcohol Abuse and Alcoholism. (2004). *NIAAA council approves definition of binge drinking*. NIAAA Newsletter: Winter.

Newcomb, M. D., & Felix-Ortiz, M. (1992). Multiple protective and risk factors for drug use and abuse: Cross-sectional and prospective findings. *Journal of Personality and Social Psychology, 63*(2), 280–296.

Nichols, S. L., Lowe, A., Zhang, X., Garvie, P. A., Thornton, S., Goldberger, B. A., et al. (2014). Concordance between self-reported substance use and toxicology among HIV-infected and uninfected at risk youth. *Drug and Alcohol Dependence, 134,* 376–382.

Nishimura, S. T., Hishinuma, E. S., Miyamoto, R. H., Goebert, D. A., Johnson, R. C., Yuen, N. Y., et al. (2001). Prediction of DISC substance abuse and dependency for ethnically diverse adolescents. *Journal of Substance Abuse, 13*(4), 597–607.

Oetting, E. R., Beauvais, F., & Edwards, R. (1999). *The American drug and alcohol survey*. Fort Collins, CO: Rocky Mountain Behavioral Science Institute Inc.

Peleg, K., Gopher, A., Jaffe, D. H., Siman-Tov, M., & Almog, S. (2010). Comparison of blood alcohol levels with breath alcohol levels measured using the drager 7110 MKIII breathalyzer. *Injury Prevention, 16*(Suppl 1), A147–A148.

Rahdert, E. R. (1991). *The adolescent assessment/referral system manual*. Washington, DC: Westover Consultants Inc.

Read, J. P., Kahler, C. W., Strong, D. R., & Colder, C. R. (2006). Development and preliminary validation of the young adult alcohol consequences questionnaire. *Journal of Studies on Alcohol and Drugs, 67*(1), 169–177.

Reinert, D. F., & Allen, J. P. (2007). The alcohol use disorders identification test: An update of research findings. *Alcoholism, Clinical and Experimental Research, 31*(2), 185–199.

Riggs, S. R., Alario, A., Dube, C., Goldstein, M., Lewis, D., Myers, E., et al. (1989). *Adolescent substance use instructor's guide. The project ADEPT curriculum for primary physician training*. Providence, RI: Brown University.

Robinson, S. M., Sobell, L. C., Sobell, M. B., & Leo, G. I. (2012). Reliability of the timeline followback for cocaine, cannabis, and cigarette use. *Psychology of Addictive Behaviors, 28*(1), 154–162.

Rogers, R., Cashel, M. L., Johansen, J., Sewell, K. W., & Gonzalez, C. (1997). Evaluation of adolescent offenders with substance abuse: Validation of the SASSI with conduct-disordered youth. *Criminal Justice and Behavior, 24*(1), 114–128.

Rosen, C. S., Henson, B. R., Finney, J. W., & Moos, R. H. (2000). Consistency of self-administered and interview-based addiction severity index composite scores. *Addiction, 95*(3), 419–425.

Rounsaville, B. J., Kosten, T. R., Weissman, M. M., & Kleber, H. D. (1986). Prognostic significance of psychopathology in treated opiate addicts: A 2.5-year follow-up study. *Archives of General Psychiatry, 43*(8), 739–745.

Rychtarik, R. G., Koutsky, J. R., & Miller, W. R. (1998). Profiles of the alcohol use inventory: A large sample cluster analysis conducted with split-sample replication rules. *Psychological Assessment, 10*(2), 107–119.

Rychtarik, R. G., Koutsky, J. R., & Miller, W. R. (1999). Profiles of the alcohol use inventory: Correction to Rychtarik, Koutsky, and Miller (1998). *Psychological Assessment, 11*(3), 396–402.

Saitz, R., Cheng, D. M., Allensworth-Davies, D., Winter, M. R., & Smith, P. C. (2014). The ability of single screening questions for unhealthy and other drug use to identify substance dependence in primary care. *Journal of Studies on Alcohol and Drugs, 75,* 153–157.

Samet, S., Waxman, R., Hatzenbuehler, M., & Hasin, D. S. (2007). Assessing addiction: Concepts and instruments. *Addiction Science & Clinical Practice, 4*(1), 19–31.

Saunders, J. B., Aasland, O. G., Babor, T. F., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*(6), 791–804.

Schottenfeld, R. S., & Pantalon, M. V. (1994). Assessment of the patient. In M. Galanter & H. D. Kleber (Eds.), *Textbook of substance abuse treatment*. Washington, DC: American Psychiatric Press.

Schwartz, R. H., & Wirtz, P. W. (1990). Potential substance abuse detection among adolescent patients using the drug and alcohol problem (DAP) quick screen, A 30-item questionnaire. *Clinical Pediatrics, 29*(1), 38–43.

Selzer, M. L. (1971). The Michigan alcoholism screening test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127*(12), 1653–1658.

Selzer, M. L., Vinokur, A., & Rooijen, L. V. (1975). A self-administered short Michigan alcoholism screening test (SMAST). *Journal of Studies on Alcohol and Drugs, 36*(01), 117–128.

Shamblen, S. R., & Dwivedi, P. (2010). Is some data better than no data at all? Evaluating the utility of secondary needs assessment data. *Drugs: Education, Prevention and Policy, 17*(6), 835–852.

Simons, J. S., Dvorak, R. D., Merrill, J. E., & Read, J. P. (2012). Dimensions and severity of marijuana consequences: Development and validation of the Marijuana consequences questionnaire (MACQ). *Addictive Behaviors, 37*(5), 613–621.

Skinner, H. A. (1982). The drug abuse screening test. *Addictive Behaviors, 7*(4), 363–371.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow–back: A technique for assessing self–reported alcohol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biochemical methods*. Totowa, NJ: Humana Press.

Stein, L. A. R., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Using the MMPI-2 to detect

substance abuse in an outpatient mental health setting. *Psychological Assessment, 11*(1), 94–100.

Stein, L. A., Lebeau, R., Clair, M., Rossi, J. S., Martin, R. M., & Golembeske, C. (2010). Validation of a measure to assess alcohol-and marijuana-related risks and consequences among incarcerated adolescents. *Drug and Alcohol Dependence, 109*(1), 104–113.

Stephens, R. S., Roffman, R. A., & Curtin, L. (2000). Comparison of extended versus brief treatments for marijuana use. *Journal of Consulting and Clinical Psychology, 68*(5), 898–908.

Stone, A. L., & Latimer, W. W. (2005). Adolescent substance use assessment: Concordance between tools using self-administered and interview formats. *Substance Use and Misuse, 40*(12), 1865–1874.

Substance Abuse and Mental Health Services Administration. (2002). Binge drinking among underaged persons. *The NHSDA Report.* April 11. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration. (2013). *Results from the 2012 national survey on drug use and health: Summary of national findings*. NSDUH series H-46, HHS Publication No. (SMA) 13-4795. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Summerfeldt, L. J., & Antony, M. M. (2002). Structured and semistructured diagnostic interviews. In A. M. Antony (Ed.), *Handbook of assessment and treatment planning for psychological disorders*. New York, NY: Guilford Press.

Sussman, S. Y., & Ames, S. L. (2001). *The social psychology of drug abuse*. Buckingham, UK: Open University Press.

Sussman, S., & Ames, S. L. (2008). *Drug abuse: Concepts, prevention, and cessation*. New York, NY: Cambridge University Press.

Svanum, S., Levitt, E. E., & McAdoo, W. G. (1982). Differentiating male and female alcoholics from psychiatric outpatients: The MacAndrew and Rosenberg alcoholism scales. *Journal of Personality Assessment, 46*(1), 81–84.

Tarter, R. E. (1990). Evaluation and treatment of adolescent substance abuse: A decision tree method. *The American Journal of Drug and Alcohol Abuse, 16*(1–2), 1–46.

Tarter, R. E., & Hegedus, A. M. (1991). The drug use screening inventory: Its applications in the evaluation and treatment of alcohol and other drug abuse. *Alcohol Health & Research World, 15*(1), 65–75.

Tarter, R. E., Laird, S. B., Bukstein, O. G., & Kaminer, Y. (1992). Validation of the adolescent drug use screening inventory: Preliminary findings. *Psychology of Addictive Behaviors, 6*(4), 233–236.

Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics, 42*(5), 441–447.

Tonigan, J. S., & Miller, W. R. (2002). The inventory of drug use consequences (InDUC): Test-retest stability and sensitivity to detect change. *Psychology of Addictive Behaviors, 16*(2), 165–168.

Uhl, G. R., Drgon, T., Johnson, C., Fatusin, O. O., Liu, Q. R., Contoreggi, C., et al. (2008). "Higher order" addiction molecular genetics: Convergent data from genome-wide association in humans and mice. *Biochemical Pharmacology, 75*(1), 98–111.

Vartanian, T. P. (2010). *Secondary data analysis*. New York, NY: Oxford University Press.

Vincent, E. C., Zebelman, A., & Goodwin, C. (2006). What common substances can cause false positives on urine screens for drugs of abuse? *Clinical Inquiries, 55*(10), 893–894.

Voas, R. B., Furr-Holden, C. D. M., Lauer, E., Bright, C., Johnson, M. B., & Miller, B. (2006). Portal surveys of timeout drinking locations: A tool for studying binge drinking and AOD use. *Evaluation Review, 30*(1), 44–65.

Vrieze, S. I., Iacono, W. G., & McGue, M. (2012). Confluence of genes, environment, development, and behavior in a post-GWAS world. *Development and Psychopathology, 24*(4), 1195–1214.

Vrieze, S. I., McGue, M., Miller, M. B., Hicks, B. M., & Iacono, W. G. (2013). Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: Twin biometry, GCTA, and genome-wide scoring. *Behavior Genetics, 43*(2), 97–107.

Weed, N. C., Butcher, J. N., McKenna, T., & Ben-Porath, Y. S. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2: The APS and AAS. *Journal of Personality Assessment, 58*(2), 389–404.

Wertz, J. S., Cleaveland, B. L., & Stephens, R. S. (1995). Problems in the application of the addiction severity index (ASI) in rural substance abuse services. *Journal of Substance Abuse, 7,* 175–188.

White, H. R., & Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol and Drugs, 50*(01), 30–37.

Winters, K. C. (1999). Treating adolescents with substance use disorders: An overview of practice issues and treatment outcome. *Substance Abuse, 20*(4), 203–225.

Winters, K. C., & Henly, G. A. (1989). *Personal experience inventory (PEI) test and manual*. Los Angeles, CA: Western Psychological Services.

Winters, K., & Henly, G. (1993). *Adolescent diagnostic interview manual (ADI)*. Los Angeles, CA: Western Psychological Services.

Winters, K. C., Stinchfield, R. D., & Henly, G. A. (1993). Further validation of new scales measuring adolescent alcohol and other drug abuse. *Journal of Studies on Alcohol and Drugs, 54*(5), 534–541.

Yudko, E., Lozhkina, O., & Fouts, A. (2007). A comprehensive review of the psychometric properties of the drug abuse screening test. *Journal of Substance Abuse Treatment, 32*(2), 189–198.

Zanis, D. A., McLellan, A. T., & Randall, M. (1994). Can you trust patient self-reports of drug use during treatment? *Drug Alcohol Dependence, 35*(2), 127–132.

Validity: Conceptual and Methodological Issues in Substance Abuse Research

12

Brian E. Perron, David Cordova, Christopher Salas-Wright, and Michael G. Vaughn

12.1 Introduction

> Accurate and minute measurement seems to the nonscientific imagination, a less lofty and dignified work than looking for something new. But nearly all the grandest discoveries of science have been but the rewards of accurate measurement and patient long-continued labour in the minute sifting of numerical results.
>
> —Baron William Thomson Kelvin

Although Kelvin made these remarks nearly 150 years ago in his presidential address to the British Association for the Advancement of Science (1871), they remain relevant to our current scientific initiatives, including those in substance abuse research. In fact, it is difficult to identify a single advancement in substance abuse research that was not due, in part, to the advancement of measurement. Measurement is the process of assigning a numeric value to, or ordering characteristics or traits of entities under investigation, and it represents one of the most important and difficult tasks in science. The quality of knowledge gained from any scientific investigation is largely dependent upon the quality of its measurement.

In this chapter, we focus our attention on one particular aspect of measurement in substance abuse research: validity. As such, there are several reasons for having chosen this focus. First, validity is the most important issue among all measurement issues, so it necessarily deserves critical and careful attention in any discussion of measurement. Furthermore, in the absence of a strong foundation of validity evidence, research is meaningless. Second, validity can be easily misunderstood, as it is a complex, continually evolving topic. Some discussions of validity in substance abuse research are disconnected from modern validity theory, and the subsequent content aims to address this problem. Third, a number of strategies for gathering validity evidence related to our measurements are underutilized in substance abuse research, maintaining instead a strong reliance on analytic strategies that may be driven more by precedent than validity theory itself; once again, this as an opportunity to identify such shortcomings and make suggestions for improvements.

The following framework first provides a definition of validity, highlighting the major differences between modern and traditional validity theory. Attention is then diverted to the methodological and conceptual issues associated with gathering validity evidence in substance abuse research. This discussion focuses primarily

B.E. Perron (✉) · D. Cordova
School of Social Work, University of Michigan, Ann Arbor, MI, USA
e-mail: beperron@umich.edu

D. Cordova
e-mail: beperron@umich.edu

C. Salas-Wright
School of Social Work, The University of Texas at Austin, Austin, TX, USA
e-mail: cpsw@bu.edu

M.G. Vaughn
School of Social Work,
Saint Louis University, Saint Louis, MO, USA
e-mail: mvaughn9@slu.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_12

on self-report measures, given that these types of measures are among the most common data collection strategies in substance abuse research, and often give rise to a myriad of problems that threaten the measurement validity.

It should be noted that it is beyond the scope of this chapter to provide a comprehensive review of conceptual and methodological issues associated with validity in substance abuse research. However, carefully selected readings will provide additional details or excellent examples for sources of validity. These readings are presented as an Appendix to the chapter, and the reader is encouraged to use these resources for additional and more in-depth coverage of the various sources of validity evidence covered in this chapter.

12.2 What Is Validity?

Many definitions of validity can be found in substance abuse research. Most of them, in one way or another, define validity as the extent to which a test measures what it purports to measure. Traditionally, we often hear about 'types' of validity, such as construct validity, content validity, convergent and divergent validity, face validity, and factorial validity, and it is fairly common for substance abuse researchers to make statements about an instrument or scale being a *valid* measure of a given construct. Such statements are often supported by studies that have employed factor analysis or other correlational strategies to establish a strong profile of the study's psychometric properties.

This understanding of validity and approach to validation is not, however, fully consistent with *modern validity theory*. Modern validity theory is associated with the highly influential work of Messick (1989, 1995), who defined validity as "An integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (1989). Messick was highly influential in shaping the *Standards for Educational and Psychological Testing*, which is referred to hereafter as the *Standards*; (American Educational Research Association et al. 1999). The *Standards* was authored jointly by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, and it is widely regarded as the gold standard for measurement definitions. The *Standards* presents a definition that is much less technical, yet still aligns with the work of Messick: "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association et al. 1999).

One important distinction of modern validity theory, as compared to earlier views of validity theory (hereafter referred to as 'traditional validation theory'), is that validity is *not* a property of measurement. From the perspective of modern validity theory, it is technically incorrect to state that a measure is (or is not) valid, such as 'This is a valid measure of alcohol expectancies.' In modern validity theory, validity refers to the extent to which the interpretations and uses of scores are consistent with evidence and theory. A measure might have strong validity evidence to support its use in one situation but not another. For example, one might choose to use an alcohol expectancy measure as an indicator of risky drinking, even though the evidence and theory would not support such a use. This, however, does not mean that the measure itself is *invalid,* but rather, the validity issue is concerned with the measurements and how they are interpreted and used.

Another way that modern validity theory differs from traditional understanding is that modern validity theory treats validity as a unitary concept that is not divided into various *types*. From the modern perspective, construct validity (or face validity, content validity, etc.) is simply validity. However, evidence of validity can be obtained from different sources, and these sources correspond to traditional conceptualizations of validity types (Table 12.1). These sources of evidence should then be integrated with theory in order to establish an argument for validity. Ultimately, an argument for validity should not focus

Table 12.1 Sources of validity evidence from modern validity theory

Evidence source	Definition	Evidence gathering strategies
Instrument content	Demonstrating that elements of an instrument are relevant to and representative of the target construct	(a) Conducting a conceptual analysis to establish the logical links between the words and phrases or observations used in an instrument and the construct being measured, and (b) Quantifying judgments about the elements of an instrument using formalized scale procedures with multiple experts
Response process	– The fit between the construct and the detailed nature of performance or response actually engaged in by examinees (American Educational Research Association et al. 1999) – A measure of self-esteem should not be influenced by social desirability	(a) Using think-aloud protocols to understand how individuals are responding to or interpreting items on an instrument, and (b) Applying statistical techniques to test for response sets, such as an acquiescence bias
Internal structure	Showing that the items and subscales exhibit patterns of association consistent with substantive theory	(a) Using estimates of internal consistency, and (b) Formal tests of the factor structure using confirmatory factor analysis
Associations with other variables	– Establishing the extent to which a measure agrees with another measure of the same construct (criterion evidence) – Making an interpretation of validity on the basis of empirical associations that are expected and the absence of associations that are not expected is referred to as convergent and discriminant (or divergent) evidence, respectively – Documenting the extent to which instrument-criterion evidence can be generalized to a new situation without further study of validity in that new situation (American Educational Research Association et al. 1999)	(a) Criterion evidence: This involves correlating a measure of the focal construct with a criterion that has been measured at the same point in time (concurrent), with a future criterion (predictive), or a previously measured construct (postdictive). A predictive criterion involves correlating a current measure of the focal construct with a future outcome. A postdictive criterion is similar to a predictive criterion, except that it involves use of a previously measured criterion (b) Discriminant convergent evidence: Testing hypotheses using correlation procedures

Note Table definitions, examples, and strategies adapted from Messick (1995) and the Standards (American Educational Research Association et al. 1999)

on whether or not a specific type of validity evidence was obtained; rather, it should focus on the degree of evidentiary support.

Before discussing different sources of validity evidence and their respective data gathering strategies, it is worthwhile to mention the special relationship that validity has with reliability. Reliability is the degree to which measurements are free from error, making reliability inversely related to error. Reliability goes hand-in-hand with validity; it is a necessary but not sufficient condition for validity. Reliability can exist without validity, but validity cannot exist without reliability.

Reliability is often given more attention that validity in substance abuse research. This is likely due to the fact that reliability is more amenable to quantification than validity, and many off-the-shelf software packages have built-in functions that allow complex reliability analyses to be carried out with relative ease. Although reliability is a critical issue to consider, it is most beneficial to consider all issues of reliability within the context of validity. In other words, it is important to consider how any problems with reliability may impact a study's use and interpretation of the scores, i.e., the study's validity.

12.3 Instrument Content

The content of an instrument, also referred to as its *elements,* consists of the specific words used to form the items or indicators (questions, statements, perceptions, opinions) of an instrument, as well as the formatting of individual items, response options, and guidelines for administration and scoring (American Educational Research Association et al. 1999). Validity evidence that is derived from measure content requires demonstrating that its elements are "relevant to and representative of the target construct for a particular assessment purpose" (Haynes et al. 1995). Evaluations of validity require that the target construct be explicitly defined, and that the items or indicators with the target construct have a clear theoretical relationship.

Logical and quantified judgments are key evaluation strategies for deriving validity evidence based upon content. However, neither type of evaluation leads to a statistic or coefficient that can be directly interpreted as a validity estimate. Rather, the evaluations are based on theory and interpretive arguments. As a working example for this source of validity evidence, consider the construct of *involvement in Alcoholics Anonymous* (*AA*). The reader can assume that such a measure could be used to test the hypothesis that involvement in AA has a mediating influence on formal treatment for alcoholism (Tonigan et al. 1996).

12.3.1 Logical Evaluation

A logical evaluation involves carefully considering the relevance, specificity, and clarity of each item in relation to the conceptual definition of the target construct. When an instrument includes multidimensional constructs, it is imperative that each dimension is examined separately. The extent to which other elements of a measure influence the score must also be carefully considered. The elements of each dimension need to be examined, including response format, items, and overall appearance. Let us first consider a single-item measurement strategy for AA-involvement, which is commonly measured based on AA meeting attendance. For example, the single-item measurement strategy may be, "Did you attend an AA meeting since you completed treatment?" or, "How many AA meetings did you attend since you completed treatment?"

Logical evaluations of measures must be done in context of theory. More specifically, if our theoretical framework suggests that the mediating influence of involvement in AA is accounted for by meeting attendance, then it makes theoretical sense to use this measurement strategy. The burden in making a validity argument ultimately rests on the clear articulation of this theoretical relationship and producing evidence that is consistent with theory. However, if our theory of *involvement* includes more than just meeting attendance, such as having a sponsor or assuming some type of AA-related volunteer work, then the measure does not sufficiently reflect the construct. In other words, the key issue is whether the measure (in this example, the single-item "Did you attend an AA meeting since you completed treatment?") provides enough coverage of involvement. When viewed as a multidimensional construct (including more dimensions than just meeting attendance), it does not.

Numerous attempts have been made to examine involvement in AA using a multiple indicator strategy (Humphreys et al. 1998; Tonigan et al. 1996). Let us consider the Alcoholics Anonymous Involvement (AAI) Scale, which is a 13-item self-report that conceptualizes *involvement* using two distinct but interrelated factors: (1) attendance, and (2) involvement in AA-activities (Table 12.2; Tonigan et al. 1996).

When considering the content of a scale, such as the AAI, it is important to consider the theoretical relationship between the indicator and the target construct. Each item must be relevant to and representative of the target construct. For example, the involvement factor is largely defined by the two questions pertaining to meeting attendance. However, the clarity of the attendance construct is somewhat lost by also including the "number of steps worked" as a

Table 12.2 Summary of questions of the Alcoholics Anonymous Involvement (AAI) scale

Scale questions	Response options
(Involvement)	
• Ever attended AA meeting?	Yes/No
• Attended AA last year?	Yes/No
• Ever considered self an AA member?	Yes/No
• "90 meetings in 90 days"?	Yes/No
• Celebrated AA birthday?	Yes/No
• Had an AA sponsor?	Yes/No
• Been an AA sponsor?	Yes/No
• Ever had spiritual awakening?	Yes/No
(Attendance)	
• No. of AA steps "worked"	Count
• No. of AA meetings (past year)	Count
• No. of AA meetings (lifetime)	Count

Notes All scale questions are presented in an abridged format. Refer to original source for complete wording of items. Involvement scores are computed by first assigning 1 and 0 to Yes and No (respective), and then summing these items respectively. These values are then summed to produce an involvement score. Factors are indicated by parentheses. Items on the Attendance factor have a count-based response option. Each raw score is transformed to a decile score and then divided by 10, which produces a value ranging from 0.1 to 1.00. Decile scores are the summed to produce an Attendance score. Involvement and Attendance score can also be summed to produce a composite Involvement score

unique dimension of attendance. The authors note that this item loaded equally on both the attendance and involvement factors. The item was ultimately assigned to the attendance factor because it fits with the other count-based items on the attendance factor. A stronger argument for validity can be made when measurement decisions are theoretically driven.

Discussion thus far has focused on the theoretical specification of items and questions, but keep in mind that content, as a source of validity, refers to all the elements of the instrument. One important and often overlooked element is the scoring algorithm. The scoring algorithm is an explicit set of rules for scoring a measure. The scoring algorithm for the AAI is included as a note in Table 12.2.

The AAI uses nominal response options, such as yes/no, for a number of questions. These are assigned values of 1 and 0, respectively, and subsequently summed to produce one of the two-factor scores. This is very common practice in substance abuse research when indices are used. The use of this scoring algorithm involves important assumptions that need to be carefully considered; that is, do all these activities contribute equally and linearly in the causal process? For instance, does the underlying theory of involvement suggest that "having a sponsor" is equally as important as "attending 90 meetings in 90 days"?

The transformation of the scores for the attendance factor is also deserving of consideration (refer to scoring algorithm in Table 12.2). More specifically, these count-based scores were transformed to deciles and then divided by 100. Consequently, these are no longer absolute measurements but relative measurements; a person's score on any one of these items is dependent upon their actual response, as well as how their response compares to other individuals within the sample. In other words, these scores are converted into rankings, which necessarily have implications for interpretation.

These are just a few examples of conducting a logical evaluation of the elements of a measure;

they certainly do not represent a comprehensive review. In our discussion of the AAI scale, we want to be clear that we are not criticizing this particular scale. Rather, we are only raising questions that would be consistent with a logical evaluation, and we believe that careful and critical analysis is necessary to advance the quality of the measure.

12.3.2 Judgments of Content

So far we have discussed logical evaluations as primarily a theoretical exercise conducted by a single individual. However, science is a public activity, and the interpretations from a logical evaluation may not be consistent between individuals, as differences in experience and knowledge of theory can lead to different points of view. It would be useful to have many different people participate in the logical evaluation of a measure. However, it is difficult to coordinate and summarize multiple views. One way to make the process more manageable is to quantify judgments of content. This involves devising strategies of systematizing and quantifying judgments of content through survey methods. For example, Haynes et al. (1995) recommends subjecting every element of a measure to a 5- to 7-point evaluation scale, focusing on relevance, representativeness, specificity, and clarity. This evaluative scale, along with the measurement under study, is then distributed to experts for evaluation. Rubio et al. (2003) also recommends involving both content experts and laypersons; content experts are professionals with expertise in the field relevant to the measure, and laypersons are people for whom the topic of the measure is most salient.

Data can be summarized in different ways to formulate what is referred to as a *Content Validity Index* (Davis 1992; Rubio et al. 2003). However, no guidelines or formal criteria are available to either compute the index or use such an index in making an interpretation of validity. Furthermore, it is important to keep in mind that the constructed evaluative scale is, indeed, a measurement device that is subject to the same validity concerns as the measure being evaluated. Thus, readers need to be cautious in interpreting these evaluations. Readers are encouraged to review the work of Haynes et al. (1995) for a comprehensive review of the conceptual and empirical procedures involved with using content to establish a validity interpretation for a measure.

12.4 Response Processes

Part of building a strong validity argument involves collecting evidence on the responses that reflect differences in actions, strategies, and thought processes of individual respondents (or observers). Our confidence in a measure is compromised when different categories of people reveal differences in their response patterns due to something about the measure or the measurement process. For example, when we ask somebody to *estimate* the number of drinks they had within a given timeframe, can we be sure that everybody engages in the same mental processes? Midanik and Hines (1991) examined how such standardized items frequently used in alcohol research on consumption were not always consistently understood. This revelation was made using a protocol analysis, or more commonly referred to as *think-aloud protocols*, to assess recall strategies used in responding to standard alcohol consumption items. The results of the study were surprising: subjects used a number of different recall strategies in responding to different alcohol consumption items (Midanik and Hines 1991). These findings provided compelling evidence to suggest that recall strategies respondents are using to answer standard alcohol items "may not necessarily lead to the response … that is expected by the researcher," and that "the 'cues' that are given [to] respondents by interviewers, e.g., response cards with answer categories, may not be adequate enough to elicit the type of information needed" (Midanik and Hines 1991).

We can also make incorrect inferences when we apply culturally specific descriptions and symptomatologies to other cultures (Room et al.

1996). For example, and as described by Room et al. (1996), the term "alcoholism" was defined in the 1940s and 1950s using a singular and culturally specific terminology from Alcoholics Anonymous. However, in 1960, Jellinek reported various differences in cultural expressions of alcoholism that were not fully compatible with the AA definition (1960). That is, alcoholism is largely defined by the outcomes associated with drinking, and societies have different ways for responding to drinking-related behaviors. Thus, identifying possible culturally bound definitions in our measurements is crucial in establishing a strong foundation of validity.

In the following section, we will describe the use of both qualitative and quantitative strategies for examining response processes. It should be noted that these examples reflect a number of different ways that response processes can be assessed in substance abuse research. However, keep in mind that these examples will not produce unambiguous findings; rather, they provide a new set of data that still must be interpreted in the context of theory.

12.4.1 Think-Aloud Protocols

Think-aloud protocols are a research strategy that asks participants to verbalize their thoughts while performing either a cognitive or physical task. The objective is to elicit information about the task being performed. For example, if we wanted to understand how persons engaged in the process of "estimating" the number of drinks within a two-week period, we would ask that person to *think aloud* as they are performing this mental calculation. Gardner and Tang (2013) used the following prompt in a think-aloud study of persons who were completing habit measures applied to health-related behaviors, including alcohol consumption:

> We want to examine how you interpret questions commonly used in health-related research studies. We are going to ask you to fill in a questionnaire and 'think aloud' as you fill it in. What we mean by 'think aloud' is that we want you to say everything you are thinking, from the time you first see each question until you reach a decision on how to answer the question […] as if you were alone in the room speaking to yourself.

The task of thinking out loud should not be interrupted until the individual provides the response to the stimulus question. As noted by Hevey (2010), many respondents engaged in think-aloud protocols have a tendency to explain or justify their thoughts. In doing so, they are relying on other cognitive processes that are unrelated to the task at hand, so it is important to consider strategies that minimize this possibility. For example, they may complete some simple warm-up tasks to become familiar with the process. The researcher may also consider sitting behind the respondent to further avoid social interactions, while still prompting the individual to speak aloud if there is a period of silence (Hevey 2010).

Think-aloud protocols are typically recorded and transcribed. This allows a coding strategy to be developed that is specific to the research question at hand. For example, Midanik and Hines (1991) developed a coding strategy that reflected the different cognitive strategies that respondents used to compute the number of drinks consumed within a given period. A coding strategy could also be developed to better understand how respondents are interpreting specific words and phrases. Keep in mind that any type of think-aloud protocol analysis requires careful attention to validity issues, even though the research is qualitative in nature.

12.4.2 Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is an advanced statistical procedure that is used to study the dimensionality of a set of variables. This important tool is commonly used for examining the internal structure of a measure, which will be covered in the following section. CFA is typically performed on a scale where the

items serve as indicators of a trait or factor in a common factor model. While CFA is commonly used for single samples, *measurement invariance* can be performed using a multiple group analysis. Measurement invariance is a statistical property of measurement that indicates whether, and the extent to which, the latent construct being examined varies across specific groups.

For example, Derringer et al. (2013) tested whether DSM-IV symptoms of substance dependence are equivalent between community and clinical samples. The DSM criteria were developed for application within clinical populations but not community samples; however, the DSM criteria are routinely used among community samples in epidemiologic surveys. Using the same criteria carries the assumption that the criteria are reflecting the same trait in all groups, which in this case is substance dependence. Therefore, it is necessary to determine whether the criteria have equivalent measurement properties in both populations, such as a rank order of symptom endorsement frequencies. Establishing equivalence is necessary when making inferences of community-based studies and comparing those findings with clinical samples.

While different strategies exist for testing measurement equivalence using CFA, Derringer et al. (2013) used a multi-group model in their analysis, which is one of the most common ways for testing measurement equivalence. Where traditional CFAs involve only one input matrix for the overall sample, a multiple group analysis involves a separate input matrix for each discrete group being compared; in this case, clinical versus community samples. Thus, ensuring the factor loadings for each criterion is the same across both groups is critical to establishing the property of measurement invariance. Derringer and colleagues found that the clinical sample endorsed more dependence symptoms, but the pattern of symptom endorsement was similar across groups. This provided evidence that the DSM-IV criteria are equally appropriate for describing substance dependence across different sampling methods.

12.5 Internal Structure

Internal structure refers to the theoretical and empirical relationships among items or questions that are used to measure a construct. Examination of internal structure involves a theoretical examination and an empirical analysis of the assumption of unidimensionality, while factor structure must be examined when we are working with multidimensional constructs. The empirical analyses rely on a variety of statistical procedures, but it is important to emphasize that the statistical procedures cannot be conducted without careful attention to theoretical assumptions. The following section discusses the process of examining unidimensionality and factor structure.

12.5.1 Unidimensionality

Constructs are assumed to have both theoretical and empirical unidimensionality. A unidimensional measure is composed of items that measure the same attribute of a specified construct (Hattie 1985). From a theoretical standpoint, any scale score that is derived from a measure must correspond, or have a relationship with, a single dimension of that construct. In other words, "homogenous items have but a single common factor among them that are related to the underlying factor … in a linear manner" (Green et al. 1977).

Conceptually complex constructs are measured with scales that contain *subscales* or factor scores. It is certainly acceptable and commonplace to have conceptually complex constructs, where a construct has multiple interrelated dimensions, or conceptually distinct subconstructs (McGrath 2005). For example, in the earlier example of the AAI scale, two separate but interrelated dimensions of involvement were observed: meeting attendance and involvement. In DSM-5, alcohol use disorder is defined as a single, unidimensional construct, meaning that all the criteria are assumed to be related to that

single factor. This can be compared with the earlier bi-factor model contained in DSM-IV, which divides the disorder into two conceptually distinct constructs: alcohol abuse and alcohol dependence.

Each dimension of a multidimensional construct must be reduced to reflect a single theoretical dimension of the construct. Unless the scale score reflects a single dimension, it is impossible to establish meaningful associations between variables, to order people based on a specified attribute, to examine individual differences, or to create groups (Hattie 1985). The unidimensionality of constructs represents one of the most basic assumptions in measurement.

Internal consistency is a statistical procedure to represent the interrelatedness among a set of items, most commonly summarized using Cronbach's alpha (Cronbach 1951). Unfortunately, this statistic is often incorrectly used as evidence for establishing item homogeneity or unidimensionality of measure. As argued by Green et al. (1977), a high measure of internal consistency will result when a general common factor is present among a set of items. However, it is also possible to obtain a high measure of internal consistency in the absence of a common factor, or among a set of heterogeneous items. This can happen because measures of internal consistency are influenced positively by the number of items, by the number of conceptually redundant or parallel items, by the number of factors pertaining to each item, and by the magnitudes of the correlations (Hattie 1985). Items can then be correlated for a number of different reasons, and only one of which is dependent on or are caused by the same underlying construct.

Hattie (1985) provides an extensive summary of different methods for examining the dimensionality of a measure, including indices based on response patterns, reliability, principal components and factor analysis, and latent trait analysis. Each index has its own unique set of strength and weaknesses, but no single index has emerged as a gold standard. Guided by theory, it must be determined that all the items are intercorrelated in a consistent direction. Intercorrelations should be examined using the general guidelines suggested by Kline (1979). As a starting point, the range of acceptable correlation magnitudes would be from 0.30 to 0.70, in order to provide some degree of homogeneity while ensuring that the coverage is not too broad or too specific. Finally, a formal test of dimensionality may involve the use of confirmatory factor analysis, preferably with tests of other competing model specifications, such as the case of a unidimensional model compared with a multidimensional model.

12.5.2 Factor Structure

Factor analysis is the most common strategy for examining factor structure in substance abuse research. Exploratory factor and confirmatory factor analysis represent the primary tools. However, exploratory factor analysis provides only weak evidence of validity, as it is a primarily data-driven procedure and not driven by theory. Although evidence of relationships can inform theory, the actual relationships need to be subjected to rigorous hypothesis testing. Exploratory analysis does not have formal hypothesis testing procedures like factor analysis, and so confirmatory factor analysis will remain the focus of discussion.

Before diving into factor analysis, however, it is important to note that principal components analysis is sometimes incorrectly used in substance abuse research as a method for examining the factor structure of a measure. Principal components analysis is a tool for reducing dimensionality of a set of variables, and it is not based on a common factors model. Thus, as reviewed in other places in the literature (Fabrigar et al. 1999), this procedure is simply not appropriate for validation.

Confirmatory factor analysis can be used to test whether a measure of a construct is consistent with the researchers' understanding of the dimensionality of the construct. To conduct confirmatory factor analysis, the researcher specifies which set of indicators (measured variables) correspond to which constructs. If the measure is unidimensional, or comprised of only

one factor, then all of the items would be set to relate directly to that single factor. For example, a CFA analysis of the DSM-5 (American Psychiatric Association 2013) formulation of alcohol dependence would involve relating each diagnostic criterion to a single factor (alcohol dependence). If the construct being examined is multidimensional, much like the DSM-IV (American Psychiatric Association 2000) bi-factor model (abuse and dependence), then each diagnostic criterion would be assigned to be associated with, or *load on*, its respective factor (abuse or dependence). In the example of the AAI scale, both factors (involvement and attendance) were assumed to be interrelated but subsumed by an overall or second-order factor. Although this second-order factor was not explicitly named, it is assumed given that a composite score could be computed by adding the other two-factor scores together. Readers are encouraged to review the work of Rindskopf and Rose (1988) and Marsh and Hocevar (1985) for more in-depth coverage of second-order and higher order factors.

The results of a confirmatory factor analysis, typically referred to as *model fit*, are used as validity evidence. Model fit statistics show the extent to which hypothesized parameters, or relationships between each indicator and the factor(s), are consistent with the data. Besides taking into account measurement error (Kline 1998), confirmatory factor analysis can be a particularly useful tool for testing competing theories that lead to alternative model formulations, such as a single-factor model compared with a two-factor model. Other types of constraints can also be imposed to further test theory. Consistency between what we expect to find (based on theory) and what is actually found (based on the data) provides evidence to either support, or fail to support, validity interpretations to some degree.

While studies involving confirmatory factor analysis typically offer interpretations based on extensive data presentations, it is imperative that researchers do not underestimate the importance of theory. Model modifications must not occur without clear theoretical explication. Model modifications can be used to inform theories, which can then be used to inform measure revisions. This is fundamentally different than revising theory to fit a measure. Model modifications can be informative, but any modifications made to the model cannot be tested with the same data because of chance variation present in any sample. In fact, chance variation can be a source of invalidity, particularly in the form of construct irrelevant variance, which is reviewed later in this chapter.

12.6 Associations with Other Variables

Evidence regarding the associations of a construct with other constructs is central to establishing its merit relative to a given theory. Three different types of associations are relevant to this form of evidence: instrument-criterion relationships, convergent–divergent evidence, and generalizability.

12.6.1 Instrument–Criterion Relationships

An interpretation of validity can be based on the extent to which a measure agrees with another measure of the same construct. This type of validity is assessed with relational strategies, and typically empirical correlations. In other words, the target construct is correlated with another measure (the *criterion*), and the statistical significance and magnitude of the correlation is the basis for making a validity interpretation. In this section, we will consider two different time-based forms of criterion measures: concurrent and predictive.

A concurrent criterion involves correlating a measure of the target construct with a criterion that has been measured at the same point in time. For example, the underlying theory of an alcohol use disorder implies that greater impairments in functioning (social, occupational, and/or educational) correspond with an increase in the severity of the disorder. This is a hypothesized

directional relationship. Aside from the positive or negative direction of the relationship, it is important to also consider the magnitude of the relationship as suggested by theory; hypothesizing the magnitude of a relationship is possible only with well-developed theory.

A predictive criterion involves correlating a current measure of the focal construct with a future outcome, such as correlating self-efficacy with relapse following treatment. This kind of evidence is collected when in order to use results from a measure to discern what happens at a later date. This is particularly useful when taking some course of action in the present to affect an outcome that will occur at the later date. For example, brief interventions for risky drinking behaviors typically involve a brief assessment; the assessment is the actual measure, and the predictive criterion is health-related consequences. Health-related consequences are an important criterion, because they can be predicted from a risky drinking behaviors measurement, and are what the intervention is meant to influence.

Making an interpretation of validity on the basis of a criterion has appeal for researchers, as it produces a single coefficient that is relatively easy to interpret. However, this appeal has to be considered within the context of its limitations. First, the inference of validity rests fundamentally on the validity evidence linked to the criterion measure. In addition, many of the constructs we study, particularly latent constructs like self-efficacy, recovery, and/or risk, lack proper criterion measures.

12.6.2 Convergent and Discriminant Evidence

Making an interpretation of validity on the basis of empirical associations that are expected and the absence of associations that are not expected is referred to as convergent and discriminant (or divergent) evidence, respectively. At times, convergent and discriminant evidence is confusingly described as being similar to a validity interpretation based on a criterion. This, however, is a mistake. Although the correlational procedures are the same, an interpretation of validity based on convergent and discriminant evidence refers to the extent to which *other* theoretically hypothesized measures are, or are not, related to the focal construct in the expected pattern of associations. As noted above, an interpretation of validity based on a criterion refers to the extent to which a measure correlates with an alternative measure of the same construct.

The use of convergent and discriminant forms of evidence is essential to a strong validity interpretation because the performance of a measure is potentially affected by the particular combination of variables comprising the theory. For example, a measure of motivation to change should be related to self-efficacy. That is, it is reasonable to assume that the constructs are interrelated; self-efficacy makes people feel more confident about the change process, and this sense of self-efficacy can create the conditions for people to feel motivated to change. This is where we have conceptual overlap between the constructs, although the overlap is not perfect, as people can be highly efficacious, or they know that they *could* change, but they are not motivated to do so, and vice versa. From a theoretical standpoint, we would expect self-efficacy and motivation to have at least a small-to-moderate strength relationship, but not a perfect relationship (1.0). A very high correlation (>0.90) suggests that the measures are tapping the same construct when the underlying theory implies a conceptual difference. In this sense, we have to be cautious in assuming that higher correlations are desirable, because they may simply be indicators of conceptual bloat.

For evidence of divergence, a construct that would not theoretically be correlated with motivation is preferred for type of treatment (AA, counseling, religious leader, primary care provider, etc.). In other words, motivation does not

have any particular causal mechanism that would lead to changes in preference for the type of treatment.

12.7 Conclusions and Future Directions

This chapter focuses on validity, which is the most important aspect of measurement in substance abuse research. Our primary goal was to highlight key conceptual and methodological issues within the context of modern validity theory. We focused on the major sources of validity evidence, which include instrument content, response processes, internal structure, and associations with other variables. One source of validity that we excluded from this discussion is the *consequences of measurement*, which involves evaluating the intended and unintended consequences of score interpretations, including both short-term and long-term consequences (Messick 1995; American Educational Research Association et al. 1999). This form of validity is considered the most controversial, as it is not entirely clear how it fits within the modern understanding of validity theory (Brennan 2006). In fact, a systematic review that included 10 years of research and over 2400 articles indicated a general absence of validity evidence from this particular source (Cizek et al. 2010).

Improving validity in substance abuse research necessarily requires careful attention to the nuances of modern validity theory. This careful attention needs to be reflected in the published research, with a greater emphasis on validity evidence. The most practical way of achieving this is to move beyond the traditional descriptions and reviews of *psychometric properties* and place a stronger emphasis on a key validity question: what is the theory and evidence to support the inference from the measures used? While individual authors have a responsibility for providing validity evidence in research reports, journal reviewers and editors should consider making validity a stronger priority in publishing standards.

Appendix

Suggested readings related to the theoretical and methodological issues for each major source of validity evidence:

Content

- Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, *20*(3), 269–274.
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104.

Response processes

- Chung, T., & Martin, C. S. (2005). What were they thinking?: Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug and Alcohol Dependence*, *80*(2), 191–200.
- Greenfield, T. K. (2000). Ways of measuring drinking patterns and the difference they make: Experience with graduated frequencies. *Journal of Substance Abuse*, *12*(1), 33–49.
- Hines, A. M. (1993). Linking qualitative and quantitative methods in cross-cultural survey research: Techniques from cognitive science. *American Journal of Community Psychology*, *21*(6), 729–746.
- Midanik, L. T., Hines, A. M., Greenfield, T. K., & Rogers, J. D. (1999). Face-to-face

versus telephone interviews: Using cognitive methods to assess alcohol survey questions. *Contemporary Drug Problems, 26*, 673.
- Cheung, M. W. L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling*, *9*(1), 55–77.
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*(3), 471–492.
- Abbey, A., McAuslan, P., Ross, L. T., & Zawacki, T. (1999). Alcohol expectancies regarding sex, aggression, and sexual vulnerability: Reliability and validity assessment. *Psychology of Addictive Behaviors*, *13* (3), 174.
- Wiers, R. W., Hoogeveen, K. J., Sergeant, J. A., & Gunning, W. B. (1997). High-and low-dose alcohol-related expectancies and the differential associations with drinking in male and female adolescents and young adults. *Addiction*, *92*(7), 871–888.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research*. Washington, DC: American Psychological Association.
- Agrawal, A., & Lynskey, M. T. (2007). Does gender contribute to heterogeneity in criteria for cannabis abuse and dependence? Results from the national epidemiological survey on alcohol and related conditions. *Drug and Alcohol Dependence*, *88*(2), 300–307.
- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291.
- Robinson, M. E., Myers, C. D., Sadler, I. J., Riley III, J. L., Kvaal, S. A., & Geisser, M. E. (1997). Bias effects in three common self-report pain assessment measures. *The Clinical Journal of Pain*, *13*(1), 74–81.
- Rogler, L. H., Mroczek, D. K., Fellows, M., & Loftus, S. T. (2001). The neglect of response bias in mental health research. *The Journal of Nervous and Mental Disease*, *189* (3), 182–187.

Internal Structure

- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78* (1), 98.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4* (3), 272.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research (JMR)*, *25*(2).
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9* (2), 139–164.
- Muthén, B. O., Grant, B., & Hasin, D. (1993). The dimensionality of alcohol abuse and dependence: Factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *Addiction*, *88*(8), 1079–1090.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120.

Associations with Other Variables

- Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test

(AUDIT). *Alcoholism: Clinical and Experimental Research*, *21*(4), 613–619.

- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, *29*(4), 432–442.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81.
- Hasin, D., Rossem, R., McCloud, S., & Endicott, J. (1997). Alcohol dependence and abuse diagnoses: Validity in community sample heavy drinkers. *Alcoholism: Clinical and Experimental Research*, *21*(2), 213–219.
- Hasin, D. S., Schuckit, M. A., Martin, C. S., Grant, B. F., Bucholz, K. K., & Helzer, J. E. (2003). The validity of DSM-IV alcohol dependence: What Do We Know and What Do We Need to Know? *Alcoholism: Clinical and Experimental Research*, *27*(2), 244–252.
- Hesselbrock, M., Babor, T. F., Hesselbrock, V., Meyer, R. E., & Workman, K. (1983). "Never believe an alcoholic"? On the validity of self-report measures of alcohol dependence and related constructs. *Substance Use & Misuse*, *18*(5), 593–609.
- Midanik, L. T. (1988). Validity of self-reported alcohol use: A literature review and assessment. *British Journal of Addiction*, *83*(9), 1019–1029.
- Stacy, A. W., Widaman, K. F., Hays, R., & DiMatteo, M. R. (1985). Validity of self-reports of alcohol and other drug use: A multitrait-multimethod assessment. *Journal of Personality and Social Psychology*, *49*(1), 219.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: Praeger.

Cizek, G., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement, 70*(5), 732–743.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, *16*(3), 297–334.

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194–197.

Derringer, J., Krueger, R. F., Dick, D. M., Agrawal, A., Bucholz, K. K., Foroud, T., et al. (2013). Measurement invariance of DSM-IV alcohol, marijuana and cocaine dependence between community-sampled and clinically overselected studies. *Addiction, 108*(10), 1767–1776.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272.

Gardner, B., & Tang, V. (2013). Reflecting on non-reflective action: An exploratory think-aloud study of self-report habit measures. *British Journal of Health Psychology, 19*(2), 258–273.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality1. *Educational and Psychological Measurement, 37*(4), 827–838.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139–164.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238.

Hevey, D. (2010). Think-aloud methods. In N. Salkind (Ed.), *Encyclopedia of research design*. Thousand Oaks, CA: SAGE Publications Inc.

Humphreys, K., Kaskutas, L. A., & Weisner, C. (1998). The Alcoholics Anonymous Affiliation Scale: development, reliability, and norms for diverse treated and untreated populations. *Alcoholism, Clinical and Experimental Research, 22*(5), 974–978.

Jellinek, E. M. (1960). *The disease concept of alcoholism*. New Haven, CT: Hillhouse.

Kline, P. (1979). *Psychometrics and psychology*. London, UK: Academic Press.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*(3), 562.

McGrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment, 85*(2), 112–124.

Midanik, L. T., & Hines, A. M. (1991). 'Unstandard' ways of answering standard questions: protocol analysis in alcohol survey research. *Drug and Alcohol Dependence, 27*(3), 245–252.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education/Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*(1), 51–67.

Room, R., Janca, A., Bennett, L. A., Schmidt, L., & Sartorius, N. (1996). WHO cross-cultural applicability research on diagnosis and assessment of substance use disorders: An overview of methods and selected results. *Addiction, 91*(2), 199–220.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94–104.

Tonigan, J. S., Connors, G. J., & Miller, W. R. (1996). Alcoholics Anonymous Involvement (AAI) scale: Reliability and norms. *Psychology of Addictive Behaviors, 10*(2), 75–80.

Using Surveys to Study Substance Use Behavior

13

Timothy P. Johnson and Jonathan B. VanGeest

13.1 Introduction

Social and epidemiologic surveys have been used to investigate substance use patterns among both adolescents and adults in the United States, as well as other countries, for more than 50 years. Although many of these surveys are of considerable quality and rigor (for example, the Monitoring the Future Survey, the National Survey of Drug Use and Health, and The Youth Risk Behavior Survey), there have also been ongoing methodological criticisms and concerns regarding the ability of these surveys to accurately portray the behaviors they seek to measure (Hochhauser 1979; Merachnik 1972; Midanik 1982; Pernanen 1974; Popham and Schmidt 1981; United States General Accounting Office 1993; Weisner et al. 1995). Confronting these concerns is important given the lack of alternative methodologies for efficiently monitoring substance use behavior within large, national and sub-national populations. The following chapter examines the evidence currently available for addressing concerns regarding the use of survey research methods in order to study substance use behavior in general population studies.

The Total Survey Error Model, first delineated by Groves (1989), can be used to evaluate methodological concerns in substance use surveys. This model identifies errors associated with survey sampling, coverage, nonresponse, measurement, and processing. Briefly, sampling errors are concerned with imprecision in the sample(s) drawn when conducting a survey. Coverage errors focus on the degree to which the frames constructed for a survey provide adequate coverage of the population to be studied. Nonresponse errors address failures to contact and/or complete interviews with all of the sampled respondents (unit nonresponse), as well as failure to obtain responses to all of the questions included in a survey instrument (item nonresponse). Errors of measurement involve failures to adequately assess the variables of interest in a survey. These include specification errors, which involve failures to correctly conceptualize survey constructs, as well as factors external to the construct being measured that nonetheless influence measurement quality. Processing errors are defects in the construction of survey data sets and/or final analytic variables, and failure to make adequate adjustments for complex sample

T.P. Johnson (✉)
Survey Research Laboratory, College of Urban Planning and Public Affairs, University of Illinois at Chicago, Chicago, IL, USA
e-mail: timj@uic.edu

J.B. VanGeest
Department of Health Policy and Management, School of Public Health, Kent State University, Kent, OH, USA
e-mail: jvangees@kent.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_13

designs and survey nonresponse when analyzing survey data.[1]

An expanded elaboration of the Total Survey Error model has been recently outlined by Lavrakas (2013). This chapter, however, assembles the available empirical evidence concerned with errors in substance use surveys utilizing the original Groves (1989) framework. The remaining sections of this chapter are organized to present and summarize the available literature concerned with errors in substance use surveys across the five general types of survey error identified by the original Total Survey Error model.

13.2 Coverage Errors

Coverage errors are often a consequence of using a survey sampling frame that does not include all of the individuals in the population being studied. Although few sampling frames offer genuine 100% coverage of a population of interest, some come much closer to approximately complete coverage than do others, and it is this degree of coverage that is associated with the potential for serious error. As with all other elements of the Total Survey Error framework, coverage errors are not unique to substance use surveys. Nonetheless, as the likelihood of inclusion within a potential sampling frame can be associated with substance use behaviors, substance use research may have a unique vulnerability to coverage error.

In community surveys, there are social groups that are often systematically excluded from readily available sample frames. Examples of such groups include homeless persons, hospitalized individuals, college students living in dormitories, incarcerated persons, and military personnel living on military bases. More importantly, substance use may be particularly high within some of these non-residential populations (Crawford 1987; de Lint 1981). Some population subgroups may also use surrogate substances that are not captured by the questions included in many standardized survey instruments (Shield and Rehm 2012). Weisner et al. (1995) examined this issue by comparing prevalence estimates from a general population community survey, with data obtained from interviews with non-household populations found in several inpatient and outpatient settings, such as alcohol, drug, or mental health treatment, and criminal justice and/or welfare services; reported substance use was more common among those in such settings. For example, 11.3% of the household sample was defined as problem drinkers, as compared to the 43.1% of those in non-household agency settings. The disparities were even greater for indicators of weekly drug use (5.5% in the household sample vs. 36.5% in the agency sample) and measures of problem drinking and weekly drug use combined (2.2% in the household sample vs. 18.7% in the agency sample). Reardon et al. (2003) reported similar evidence of increased substance use and misuse among persons less likely to be sampled within single family households, as part of community-based epidemiologic surveys. Research by Mäkelä and Huhtanen (2010) examined alcohol-related deaths in Finland among persons excluded from traditional household sampling frames, providing evidence of increased alcohol-related mortality (and likely alcohol consumption) among those typically left out of survey sampling frames. They additionally observed, however, that the relatively small size of those groups excluded from survey sampling frames suggests that their systematically higher rates of alcohol use would nonetheless have only a small effect on overall estimates.

Telephone coverage has never been totally adequate for conducting substance use surveys, as several studies have documented higher rates of alcohol use in those housing units that do not have telephones (Gfroerer and Hughes 1991; Nelson et al. 2003). In recent years, the proliferation of cell phone ownership and the advent of wireless-only households that cannot be contacted via traditional landline telephones have led

[1]The Total Survey Error model has helped organize several decades of empirical research concerned with various sources of survey errors into a single unifying theoretical framework.

to additional survey coverage problems. It is now recognized that failure to incorporate cell-phone-only households into random digit dialed (RDD) telephone samples can lead to underrepresentation of young adults who are at higher risk for substance use behaviors. Delnevo et al. (2007), for example, found significantly decreased measures of binge drinking and heavy alcohol consumption, between both 2001–2003 and 2003–2005, in the national Behavioral Risk Factor Surveillance System (BRFSS) telephone surveys, which they attributed to the decreasing accessibility of young adults using traditional landline telephones. Research employing the U.S. National Health Interview Survey, which relies on face-to-face interviews, has demonstrated that adults in cell-phone-only households are more likely to report past year binge drinking behavior (37.6%), compared to those residing in households with landlines (18.0%), and to those in households with no telephone service (23.0%; Blumberg et al. 2006). The effects of excluding cell-phone-only households from survey estimates of binge drinking were found to be particularly serious for young adults (aged 18–-29 years) and low-income persons (Blumberg and Luke 2009). Similar relationships between phone subscribership type and substance use reports have been identified in Australia (Livingston et al. 2013) and in other U.S. studies (Hu et al. 2011). As rates of cell-phone-only residences continue to grow (Blumberg and Luke 2016), the coverage error associated with these exclusions will continue to increase, and it will become increasingly difficult to produce representative prevalence estimates using traditional landline-only sample frames. In the future, telephone surveys will likely be conducted using frames of cell phone numbers only, potentially leading to additional concerns regarding adequate coverage for substance use surveys.

Coverage errors are also common in school-based surveys, as substance use (Chavez et al. 1989; Hall 2014; Swaim et al. 1997) and earlier age of substance use onset (Ellickson et al. 1998; Horwood et al. 2010) are both known to be associated with dropping out of school. Consequently, surveys of adolescents that are school-based may often underestimate substance use, though it is important to note that many school-based surveys do not make generalizations to nonschool populations. A recent analysis by Gfroerer et al. (2012) using pooled data from the 2002 to 2008 NSDUH (National Survey of Drug Use and Health, formerly known as the National Household Survey of Drug Abuse, or NSHDA) surveys found that substance use estimates were higher for most substances among school dropouts, as compared to their same-aged counterparts in school. The effects of dropout on overall estimates increased from the 8th to the 12th grades as the numbers of dropouts increased. At the 12th grade level, they found that failure to account for dropouts would miss more than half of past year cocaine users, more than half of all lifetime Ecstasy users, 30% of current binge alcohol users, and 25% of current alcohol users.

School absenteeism is also associated with increased substance use (Bachman et al. 1981; Cowan 2001; Guttmacher et al. 2002; Kandel 1975). Given this fact, Gfroerer et al. (2012) investigated the effects of school absenteeism on substance use prevalence estimates in the NSDUH. They reported that those students who missed more days of school were also more likely to be current alcohol users, binge drinkers, and/or marijuana users. In recognition of this problem, some surveys, such as the YRBS, conduct "make-up" sessions to maximize student opportunities to participate and minimize coverage errors. It is also important to note that school policies that punish substance using students with temporary suspension or permanent banishment may also contribute to under-coverage problems

Coverage errors in substance abuse surveys may be particularly problematic with specific populations of interest, including racial and ethnic and sexual minority populations (Andersen et al. 2004; Elliott et al. 2008; Kalsbeek 2003; Meyer and Wilson 2009). Depending on the

target population, some may actually qualify as rare, further exacerbating the problem (Wagner and Lee 2015).[2] Where present, internal diversity can also present unique challenges to the researcher (Kalsbeek 2003; Meyer and Wilson 2009). Other factors also play a role. For instance, with regard to sexual minority populations, sampling procedures can often require the respondent to "come out" and acknowledge their status to participate; problematic given the highly stigmatized nature of the sexual minority identity (Meyer and Wilson 2009). Methodological challenges are compounded when sampling a population consisting of multiple minority statuses (DeBlaere et al. 2010; Mays and Jackson 1991). The quality of the sample has direct implications for the quality of statistics derived from a study. Solutions, such as oversampling, can also have important negative implications for quality of estimates, including diminished cost-effectiveness based on the size of the supplement (Andersen et al. 2004; Elliott et al. 2008; Kalsbeek 2003). Despite obstacles, the need for minority participation in substance use research is highlighted by the higher rates of substance use and abuse in some of subgroups (Center for Behavioral Health Statistics and Quality 2015; Kerr and Oglesby (this volume)).

13.3 Sampling Errors

Substance use surveys commonly employ both probability and non-probability sampling methods. When probability sampling strategies are used, all elements within the sample frame have a known, albeit not necessarily equal, probability of selection. The precision of survey statistics derived from such samples can be calculated with a good degree of confidence and used to estimate the sampling errors associated with those statistics. All other things equal, the size of a random survey sample is inversely associated with the degree of potential sampling error. The precision of survey estimates also decreases as probability samples deviate from simple random sampling designs, a commonplace occurrence designed to reduce survey costs. Of all the sources of total survey error, the sampling errors related to probability-based sample designs are probably the most well understood, and definable, in practice. While ideal, the greatest disadvantage of probability sampling low-frequency populations—often characteristic of substance abuse research—remains cost. Many probability samples are subsets of larger scale community-based studies and necessarily require significant investment in screening or contacts in order to achieve a suitable sample of the subpopulations of interest.[3] More complex variations, including stratified random sampling involving division of a population into smaller strata (e.g., schools, cities, neighborhood, etc.) of known probability, do improve cost-effectiveness, but with the noted tradeoffs in precision and potential adverse statistical effects.

Non-probability samples are commonly used when research questions focus on special populations believed to be at increased risk for substance use and misuse. There are a variety of well-known non-probability, or convenience, sample designs commonly used in practice. One of the more popular approaches currently is known as respondent-driven sampling (RDS), which was developed by Heckathorn (1997, 2002) and has been used in numerous substance use studies (Bauermeister et al. 2012; McKnight et al. 2006; Ramirez-Valles et al. 2010). Other popular non-probability strategies in substance use research include venue and facility-based sampling (Clatts et al. 2005; Halkitis et al. 2005; Kassira et al. 2001; Safika et al. 2011), snowball sampling (Kaplan et al. 1987; Sharma et al.

[2]Wagner and Lee (2015) define rare populations as referring to subgroups in the total population that are either small in size or "hard-to-reach" in terms of ability to identify or interview, such as cases where the population is geographically dispersed.

[3]Low numbers of a subsample, in and of itself, affects the precision of information gathered, as smaller numbers of subjects are associated with margins of error that make it more difficult to be confident in the actual, underlying data (Andersen et al. 2004). Also problematic is that small sample sizes may preclude analyses of complex models on the causes and consequences of substance abuse within subpopulations of interest.

2002), timespace sampling (Fernández et al. 2005; Muhib et al. 2001; Valleroy et al. 2000), as well as advertising for volunteers (Barrett et al. 2005; Levy et al. 2005; McElrath 2005). An important advantage of these designs is their cost-effectiveness when researching hard-to-find populations, such as illicit drug users. As selection probabilities are unknown, however, there are no definable sampling errors associated with these designs. Rather, non-probability-based sample designs typically suffer from large coverage errors and sampling errors are completely unknown.[4]

13.4 Nonresponse Errors

Unit response rates in general population surveys have been declining for several decades in developed countries (Groves and Couper 1998; Groves et al. 2002; Steeh et al. 2001). Survey response rates have been traditionally employed as an indicator of survey quality in general and nonresponse error in particular (Johnson and Owens 2004). Recent research, however, has demonstrated that response rates are not strongly associated with nonresponse bias, at least in the available research literature (Groves and Peytcheva 2008; Keeter et al. 2000; Merkle and Edelman 2002). Rather, it is the degree to which survey respondents and nonrespondents differ from one another in terms of variables of interest, combined with the survey's response rate that defines nonresponse bias. As an example, a British study reported by Plant et al. (1980) compared two sets of survey data, one with a 25% response rate and the other with a 79% response rate, respectively. No important differences in self-reports of alcohol consumption were found between the two.

Of course, when considering substance use behaviors, there are valid reasons to be concerned about differences between survey respondents and nonrespondents. Pernanen (1974) has suggested that persons who drink heavily might be more difficult to contact as part of survey efforts, as well as less likely to cooperate when contacted. De Lint (1981) reported in a Canadian survey that more in-person contact attempts were required to interview respondents who reported greater frequency of alcoholic beverage purchases. In addition, Cottler et al. (1987) found that respondents diagnosed with alcohol abuse and dependence required a greater number of contact attempts in order to complete interviews. Crawford (1987) additionally reported higher levels of alcohol consumption among the respondents most difficult to contact. Using a Swedish population register, Tibblin (1965) found higher rates of survey non-participation among middle-aged men who were known to have experienced alcohol-related problems. There is also some general evidence that survey nonresponse is greater among persons with poor health (Cohen and Duffy 2002; Hoeymans et al. 1998; Tolonen et al. 2010).

Another Swedish study reported that survey respondents were less likely to have been hospitalized with alcohol diagnoses, as compared to nonrespondents (Romelsjö 1989). These findings have been interpreted as evidence that heavy drinking may be a barrier to participation in social surveys due to difficulty in making contact and also in convincing those individuals who are contacted to agree to participate (Lahaut et al. 2002). Other investigations, however, have reported no differences in alcohol use between those who do and do not participate in epidemiologic surveys (Caspar 1992; Gmel 2000; Iversen and Klausen 1986; Lemmens et al. 1988; Macera et al. 1990), and, in a few cases, alcohol nonusers have also been found to be underrepresented (Dunne et al. 1997; Lahaut et al 2002; van Loon et al. 2003).

It is also important to recognize that standard field procedures in many surveys deliberately exclude active substance users from participation. Much research explicitly requires interviewers to not conduct interviews with individuals who are visibly intoxicated or who appear to be high on other substances. Kish (1965) commented on this problem nearly 50 years ago, referencing a case

[4]See the chapter by Gfroerer et al. (this volume) for additional detail regarding sample designs in substance use research.

in which a respondent was drunk by the time they came home after work every day throughout a survey's field period. While such protocols are necessary for orderly data collection and may be invoked only rarely in practice, the potential effects of such protocols on nonresponse bias must be considered. In addition, despite some claims to the contrary (Zhao et al. 2009), knowledge that a survey is concerned with substance use appears to have no effect on respondent willingness to participate (Plant and Miller 1977; Romelsjö 1989).

Research into correlates of attrition in panel surveys, in which the same respondents are interviewed at multiple time points, also provide relevant information. Several of these studies have documented higher levels of attrition among high alcohol and drug users (Bailey et al. 1992; Beard et al. 1994; Bucholz et al. 1996; Caetano et al. 2003; Goldberg et al. 2006; Hansen et al. 1985; McCabe and West 2016; McCoy et al. 2009; Paschall and Freisthler 2003; Snow et al. 1992; Wild et al. 2001). Attrition among U.S. military personnel has been associated with tobacco (but not alcohol) use, and drug use among those with less than a college education (Cunradi et al. 2005). Other research, however, has found higher attrition among nonusers (Dawson et al. 2014; Garcia et al. 2005), and additional studies have found both high alcohol intake and abstinence to be associated with increased likelihood of panel attrition (Thygesen et al. 2008; Torvik et al. 2012). Attrition was also predictive of increased mortality by way of cirrhosis of the liver and other alcoholic liver diseases. Other research has been unable to identify differences between those who do and do not drop out of panel studies, with respect to indicators of alcohol consumption and/or alcohol-related disorders (Bijl et al. 1998; de Graf et al. 2013; Lamers et al. 2012; Psaty et al. 1994).

Useful evidence also comes from research specifically designed to assess nonresponse bias. Several types of nonresponse bias studies are routinely conducted. One approach employs follow-up surveys, which typically involve obtaining survey data from nonrespondents to the primary survey (Crawford 1986). Caspar (1992), for example, conducted follow-up face-to-face interviews with a sample of nonrespondents to the 1990 NHSDA, concluding that initial nonrespondents were more likely to report lifetime drug use. Lahaut et al. (2002) provide an example of a nonresponse follow-up survey with individuals who initially did not respond to a mail survey and who were subsequently visited by interviewers to complete a face-to-face interview. Their findings suggest that abstainers were underrepresented in the initial survey. Hill et al. (1997) report a telephone follow-up survey of nonrespondents to a primary mail survey. They also found lower reporting of unsafe alcohol consumption among initial nonresponders. Lemmens et al. (1988) conducted a telephone follow-up survey of nonrespondents to a face-to-face survey, concluding that there were only small effects of nonresponse on self-reporting of alcohol consumption. A potential limitation when interpreting findings from follow-up surveys such as these is that they often use different modes of data collection between the primary survey and follow-up effort, respectively. Given what is known about mode differences in reporting of substance use behaviors, it would not be surprising that a telephone follow-up to a self-administered survey might suggest that the initial survey overestimated substance use, whereas a self-administered nonresponse follow-up survey to an initial interviewer-assisted effort might suggest that it had underestimated substance use. In each case, the effects being attributed to nonresponse bias may actually be a consequence of mode differences rather than systematic nonresponse. There are several examples in the literature of surveys that relied on interviewer-assisted follow-up interviews (cf., Hill et al. 1997; Lahaut et al. 2002) that suggest primary survey respondents over-report substance use behaviors.

Another type of nonresponse bias analysis that focuses on respondent substance use patterns are studies that compare early vs. late respondents (Lahaut et al. 2002, 2003; Trinkoff and Storr 1997; Wilson 1981). For example, Zhao et al. (2009) compared self-reports of persons

responding early and late to the Canadian Addictions Survey. Late respondents were more likely to report lifetime and past 12-month alcohol use, both chronic and acute risky alcohol use, and lifetime and past 12-month illicit drug use. Other studies have also found evidence that late survey responders are more likely to be substance users (Korkeila et al. 2001; Kypri et al. 2004, 2011). Studies such as these employ a continuum of resistance framework that assumes respondents who require greater effort to contact and interview are more similar to nonrespondents than are those who initially agree to survey requests (Lin and Schaeffer 1995; Meiklejohn et al. 2012). Other strategies compare estimates from multiple surveys (Zhao et al. 2009), compare frame data for respondents, nonrespondents, and the full sample (Lemmens et al. 1988), or compare estimates from surveys that have high versus low response rates (Plant et al. 1980).

One other nonresponse bias assessment strategy is to supplement survey data with information obtained from other sources, such as administrative records. Gfroerer et al. (1997a) examined response patterns in the 1990 National Household Survey on Drug Abuse by merging survey findings with records from the 1990 Decennial Census. Not surprisingly, this required special authorization from the government, given the strict data protections associated with the Census. These researchers found that persons with some characteristics known to be associated with substance use (such as living in urban areas and/or being male) had lower response rates, and that persons with other characteristics believed to be associated with non-substance use (older age and higher income levels) also had lower response rates. They concluded that these various nonresponse correlates would likely cancel out much of the bias either set might have introduced into the survey estimates. Research in Scotland (Gorman et al. 2014) linked data from respondents to several health surveys with death and hospitalization records, and concluded that there is an increased risk of death from alcohol-related causes among nonrespondents. Conducting nonresponse bias assessments using administrative records is common in Scandinavian nations that maintain population registers. Larsen et al. (2012) report on one study in which register data on cause-specific mortality was linked with survey nonresponse data. Fifteen-year mortality rates from alcohol-related diseases were found to be significantly greater among survey nonrespondents, as compared to respondents. In Finland, a 28-year follow-up of nonrespondents using a national mortality register revealed similar findings, as nonrespondents were several times more likely to have passed due to alcohol-related causes (Jousilahti et al. 2005).

It is also important to recognize that high nonresponse rates to individual survey questions (item nonresponse) may also be an indicator of data quality problems in substance use surveys. Some research suggests demographic variability in nonresponse rates to substance use questions. Owens et al. (2001) found that African Americans and persons who were separated or divorced were less likely, and females and persons aged 55 and older were more likely, to answer questions concerned with their use illicit drugs, respectively. Increased item nonresponse rates to substance use questions among minority groups have also been reported (Witt et al. 1992). A study by Aquilino (1992), however, reported no differences. An item nonresponse study of adolescents found higher nonresponse rates to questions concerned with alcohol and marijuana use among male respondents compared to female respondents (Stueve and O'Donnell 1997). In an item nonresponse follow-up survey conducted in the United Kingdom, Dengler (1996) found that respondents to the follow-up survey of initial nonrespondents were less likely to report drinking more than recommended amounts of alcohol, compared to initial survey respondents, leading to the conclusion that the original study had "slightly overestimated" alcohol consumption.

13.5 Measurement Errors

Measurement error occurs when survey questions fail to measure what they were designed to measure. There are several potential sources of measurement error which must be considered when constructing a survey instrument or analyzing survey data. Broadly speaking, these include design effects, respondent effects, interviewer effects, and context effects. We address each of these in the following sections.

13.5.1 Survey Design

Each element of a survey that is revealed to respondents can be expected to provide them with cues regarding the information being sought (Greenfield and Kerr 2008). These cues can be expected to influence self-reports in ways that cannot always be anticipated or controlled, and are referred to here as design-related errors. Some important design issues to be discussed include methods for asking about substance use, mode effects, use of skip patterns, and reference periods. Other design factors that may influence measurement quality include how clearly a survey is introduced to sampled respondents as being concerned with substance use, the survey's sponsor, the procedures employed to obtain respondent informed consent, the use of incentives, and the survey's focus as either primarily concerned with substance use vs. concerned with a more broad set of topics (Gfroerer et al. 2012). In regards to this later point, it has been suggested by Fowler and Stringfellow (2001) that survey respondents are more willing to discuss negative personal behaviors when they are also asked to report about positive personal behaviors and characteristics.

13.5.2 Methods for Asking About Substance Use

It will come as no surprise to most readers that the wording and structure of survey questions can be expected to have a strong influence on the answers obtained. Indeed, experimental comparisons have revealed differences in the magnitude of substance use reports obtained using various question measurement strategies. Research by Kroutil et al. (2010), for example, has demonstrated that open-ended questions seriously underestimate drug use prevalence rates, relative to close-ended questions. Similarly, Ekholm et al. (2008) reported that respondents are more likely to report binge drinking when answering close-ended questions than when answering open-ended questions.

Research has compared various methods for asking questions regarding alcohol consumption. A general finding from studies that compare various methods of collecting alcohol use reports is that asking more questions produces higher alcohol use rates (Dawson 2003; Ekholm et al. 2008; Gmel et al. 2014; Greenfield and Kerr 2008). One common measurement strategy is the quantity–frequency method, which can be used to obtain consumption information about specific types of alcohol, or which can be structured to ask about the general consumption of alcoholic beverages (often in order to save space in survey questionnaires). Serdula et al. (1999) has demonstrated that beverage-specific quantity–frequency questions produce higher estimates of alcohol consumption, compared to questions that ask more general quantity–frequency questions (a.k.a., the grouped-beverage question format). Rhem et al. (1999) have reported findings from a within-subjects experiment that documents consistently higher prevalence rates for several indicators of harmful drinking when graduated-frequency measures (Hilton 1989) are used, in comparison to the more commonly employed quantity-frequency question response format (Dawson 1998; Sobell and Sobell 2003), and weekly drinking recall questions (Rhem et al. 1999). Other studies have also found graduated-frequency measures to produce higher estimates of alcohol use in comparison to quantity-frequency measures (Midanik 1994; Poikolainen et al. 2002). The improved performance of the graduated-frequency format appears to be due to its ability to more precisely measure irregularly high levels of consumption,

although there is some evidence suggesting that the graduated-frequency approach may actually overestimate consumption (Bloomfield et al. 2013; Poikolainen et al. 2002). The usefulness of the graduated-frequency measures for cross-cultural research, however, has been called into question (Gmel et al. 2006). Newer measurement strategies, such as the Yesterday (or Recent Recall) method of reporting, in which respondents are asked to report details of their alcohol use during the previous day only, have been found to produce higher estimates than either the quantity-frequency or graduated-frequency measures (Stockwell et al. 2004, 2008). The use of a daily diary protocol for collection of alcohol consumption is frequently considered to be a "gold standard" measurement approach (Poikolainen and Kärkkäinen 1983, Poikolainen et al. 2002), but not very practical for many survey purposes.

Response category design for use in quantity and frequency questions can also influence respondent answers. For example, Schwarz (1999) has shown how simple changes in the sets of response options presented to respondents, such as emphasizing low versus high-frequency events or behaviors, can influence overall response patterns. Empirical evidence from Poikolainen and Kärkkäinen (1985) confirms this, as they obtained higher alcohol consumption reports when employing quantity and frequency questions that included heavier intake response options.

Quality-frequency measures continue to be commonly utilized in practice despite the fact that conventional wisdom among many substance use researchers holds that alcohol and drug consumption behaviors are far more variable across even brief time intervals than are assumed by these questions (Greenfield and Kerr 2008; Sobell et al. 1982). By their nature, quantity-frequency items ask for average amounts of use, eliminating the opportunity to capture episodes of heavy or binge drinking. In a community survey, Hasin and Carpenter (1998) have documented that as many as 30% of all respondents report having difficulty when answering typical survey questions concerned with usual drinking patterns due to changes in their drinking behavior during the time period in question, and that this problem is particularly acute for persons with symptoms of alcohol dependence. The key advantages of the quantity-frequency measures that make them still popular in practice are their simplicity, ease of answering, and the relatively small amount of space they require in survey instruments. Finally, the length of the recall period for which respondents are asked to report their alcohol use is also related to overall quantity of those reports (Gmel et al. 2014). In general, briefer recall windows have been demonstrated to produce higher reports of consumption (Ekholm 2004; Gmel and Daeppen 2007).[5]

13.5.3 Reference Periods

Reference periods are used to restrict and specify the time intervals for which respondents are asked to retrospectively report their substance use experiences. Commonly utilized in practice are 30-day, 12-month, and lifetime reference periods, although there are many variations. Each has their own advantages and disadvantages. It is common knowledge that recall accuracy decays with increasing length of these time intervals (Tourangeau 2000), as research suggests that greater alcohol prevalence is obtained when shorter reference periods are employed in survey questions (Bachman and O'Malley 1981; Simpura and Poikolainen 1983). Although more susceptible to recall concerns, a 12-month recall period would have the advantage of being less affected by seasonal variations in substance use (Cho et al. 2001; Greenfield and Kerr 2008). A 30-day reference period, in contrast, might be less likely to capture binge drinking episodes. Hence, many surveys ask questions about multiple reference periods in order to address the limitations of each.

[5]Comprehensive reviews by Sobell and Sobell (2003), Gmel and Rehm (2004), and Bloomfield et al. (2013) provide insight on the strengths and limitations of various approaches to measuring alcohol consumption in survey questionnaires.

Age of alcohol initiation and other drug use are topics commonly covered in substance use surveys, given that age of first substance use is considered an important risk factor for subsequent substance abuse (Grant and Dawson 1998). Unfortunately, the length of recall necessary to correctly answer these questions can be problematic for many respondents. Forward telescoping, in which respondents systematically underestimate the length of time since an event took place, is a particularly important threat to the quality of self-reports of age of first use (Shillington et al. 2012). Numerous studies have documented problems with accurate recall of this information (Bailey et al. 1992; Engels et al. 1997; Grant et al. 1995; Humphrey and Friedman 1986; Johnson and Mott 2001; Prause et al. 2007; Shillington and Clapp 2000; Shillington et al. 2011a, b, c, d).

13.5.4 Questionnaire Skip Patterns

A common question when designing substance use instruments is the issue of whether it is best to employ skip patterns, which allow respondents to avoid answering follow-up questions that are clearly not applicable to them, or to instead require all respondents to provide all answers to all items. The rationale for requiring responses to all items is twofold. First, there may be privacy concerns associated with the use of skip patterns, as those who report substance use will require more time to complete all follow-up questions, presumably allowing interviewers and/or other observers to conclude that they are in fact substance users. Second, although it is somewhat burdensome for respondents, it is likely that the presence of skip patterns will be quickly detected by many respondents and possibly motivate some to provide negative answers to filter questions in order to "skip out" of longer blocks of questions that request details regarding substance use experiences. As an example of a skip pattern, a question that asks respondents if they have ever used marijuana might be employed as a filter item. Those respondents indicating that they had used marijuana would then be eligible to answer a series of follow-up questions that queried about frequency of use, age of initiation, etc. In contrast, avoidance of skip patterns would require respondents to answer all follow-up questions, typically by selecting a "have never used marijuana" response option, which would be available for use with each follow-up question. Such an approach can considerably increase the burden and amount of time necessary to complete a questionnaire for nonusers of the substances being examined. The NSDUH has historically not employed skip patterns. An experiment reported by Gfroerer et al. (1997b) investigated the effects of using skip patterns as part of the NHSDA. In their random experiment, they found significantly lower prevalence rates for the five illicit drugs examined when skip patterns were employed. Because no differences were found in alcohol use estimates, it was concluded that privacy concerns associated with answering the most sensitive questions was a more likely explanation for the findings.

13.5.5 Mode Effects

Survey data is collected using a variety of modes, including self-administered paper-and-pencil (PAPI) and electronic questionnaires, as well as telephone and in-person interviews. Many surveys employ more than one of these modes simultaneously. The presence of mode effects in surveys has long been recognized, and there is now a considerable body of evidence documenting the effects of mode on the quality of self-reports of substance use behaviors. Most importantly, survey modes that rely on respondent self-administration have been often found to obtain greater reports of alcohol and drug use than have those modes that require respondents to directly answer questions posed by interviewers about their use of these substances (Aquilino and LoSciuto 1990; Aquilino 1994; Boniface et al. 2014; Dotinga et al. 2005; Duffy and Waterton 1984; Gfroerer and Hughes 1992; Gmel 2000; Hoyt and Chaloupka 1994; Miller et al. 2004; Romelsjö 1989; Schober et al. 1992; Tourangeau and Smith 1996; Turner et al. 1992).

There is also some evidence that these mode effects are greater for illicit substances, such as cocaine and marijuana, compared to alcohol use (Tourangeau et al. 2000).

Of the available self-administered modes, audio-computer-assisted self-interviews (ACASI) have found to generate higher reporting of substance use behaviors than have PAPI self-administered answer sheets (Brener et al. 2006; Chromy et al. 2002; Kim et al. 2008). Computer-assisted questionnaires are also known to produce data that is more internally consistent and more complete, helping to reduce the need for editing, imputation, and other processing activities that may lead to processing errors (Chromy et al. 2002; see also the discussion below in Sect. 13.5 regarding Processing Errors).

Research has, in recent years, also begun to explore the reliability and validity of substance use surveys conducted via the Internet. In a random experiment, Eaton et al. (2010) assigned classes of high school students to respond to PAPI or web questionnaires, and concluded that there were few differences in prevalence estimates obtained across the two modes. Ramo et al. (2012) examined the quality of self-reported marijuana use in a convenience sample of young adults who completed a web-based questionnaire, concluding that such data can be reliably collected. Other investigators have compared internet reporting of alcohol use with reports obtained from self-administered mail questionnaires and both face-to-face and telephone interviews, concluding that online reports have similar levels of measurement quality (Hines et al. 2010; McCabe et al. 2006; Khadjesari et al. 2009). Link and Mokdad (2005) reported obtaining higher rates of heavy drinking among respondents randomly assigned to complete a web survey, as compared to respondents completing a mail questionnaire and those interviewed by telephone. This finding is consistent with other findings suggesting higher data quality when using computer-assisted technologies.

Among interviewer-assisted survey modes, some evidence suggests face-to-face interviews appear to produce greater reports than do telephone interviews (Aquilino1992; Aquilino 1994; Gfroerer and Hughes 1991; Johnson et al. 1989), other evidence suggests no differences in substance use estimates between these two interviewer-assisted modes (Greenfield et al. 2000; Midanik and Greenfield 2003), and one study suggests higher rates of some alcohol-related measures can be obtained by telephone (Midanik et al. 2001). Some research has also reported the use of interactive voice recording (IVR) systems (a.k.a., "T-ACASI"—telephone audio computer-assisted self-interviewing) to improve the quality of substance use data collected by phone (Gribble et al. 2000; Perrine et al. 1995). In contrast to this body of findings, Bongers and van Oers (1998) found no differences in reports of alcohol use and problem drinking in a general population survey conducted in the Dutch city of Rotterdam in which respondents were randomly assigned to participate in face-to-face interviews vs. completing mail questionnaires.

13.5.6 Respondent Effects

There is considerable variability in survey respondents' ability and willingness to provide accurate answers to substance use questions. The behaviors of respondents can be understood within the framework of the cognitive model of survey response (Jabine et al. 1984), which recognizes four basic tasks required of respondents when they are answering each survey question. These include question interpretation, memory retrieval, judgment formation and response mapping, and response editing. This is a useful model for understanding how variability across respondents may influence the quality of self-reported substance use information. Evidence regarding how three of these cognitive tasks may influence the quality of substance use behavior reporting is reviewed below.

13.5.7 Question Interpretation

The risk of miscommunication may be greater in substance use surveys, compared to other topics, because respondents are known to sometimes

employ substance use terminology that differs from that employed in research questionnaires (Hubbard et al. 1992; Morral et al. 2003; Ouelett et al. 1997). The complexity of some substance use terminology may also sometimes lead to respondent confusion. This may be of greater concern in surveys of adolescents, who may not always have sufficient knowledge to correctly answer questions concerned with the use of various drugs (Chung and Martin 2005; Harris et al. 2008; Mewton et al. 2014; Morral et al. 2003; Slade et al. 2013; Swadi 1990).

Johnston and O'Malley (1997) have presented evidence suggesting that respondents sometimes are more likely to deny, or recant, having ever used certain substances that they had previously indicated having used (see also additional discussion of recanting in the section below on Response Editing). Of particular relevance here is their finding that recanting varies by type of drug being asked about, with the recanting of tranquilizers and barbiturates found to be greater than that for marijuana and cocaine, a finding that they suggest is related to the complexity of the definitions of these two substances. In alcohol research, recent reviews have additionally found that respondents commonly misinterpret standard drink sizes, suggesting that alcohol intake may be systematically underestimated in survey research (Devos-Comby and Lange 2008; Kerr and Stockwell 2012).

Another related issue is the degree to which respondent cultural background may influence the interpretation and/or comprehension of survey questions. Substance use patterns and practices are known to vary cross-culturally (De La Rosa and Adrados 1993; Room et al. 1996; Room 2007), and those varied experiences and beliefs regarding substance use can also be expected to influence respondent knowledge and familiarity with the topic in general and related terminology in particular. Many researchers recognize the importance of exploring and addressing these potential problems by employing focus groups, cognitive interviews, and ethnographic methods when designing survey instruments for use in cross-cultural settings (Gardner and Tang 2014; Midanik and Hines 1991; Ridolfo 2011; Thrasher et al. 2011).

13.5.8 Memory Retrieval

Survey methodologists have for many years been concerned with the ability and motivation of respondents to provide accurate retrospective self-reports (Friedman 1993; Sudman et al. 1996). Indeed, memory failures have been historically considered one of the more common explanations for inaccurate reporting of substance use behaviors (Pernanen 1974; Shillington et al. 2011a, b, c, d). The retrieval of the memories necessary to accurately report substance use behaviors and experiences can be particularly difficult for several reasons. Poorly worded survey questions may present respondents with difficult cognitive challenges in terms of the effort necessary to retrieve specific and/or detailed information that may not be readily accessible in memory (Wilson 1981).

As discussed earlier, questions that ask respondents to recall substance use over a longer time period tend to produce lower reporting quantities (Ekholm 2004; Gmel et al. 2014). In addition, there is evidence that heavy drinking (Babor et al. 2000; Pernanen 1974), marijuana (Solowij and Battisti 2008), cocaine (Van Gorp et al. 1999; Ardila et al. 1991; Vonmoos et al. 2013), and MDMA use (Bolla et al. 1998; Morgan 1999; Parrott et al. 1998) may be associated with impaired memory and other cognitive deficits. Mensch and Kandel (1988) have found inconsistent reporting of marijuana use to be associated with degree of drug use frequency, with more involved users providing less consistent survey responses, a finding they associate with faulty memory. Although considerable research has been invested in experimenting with strategies for aiding respondents with memory retrieval in general (Belli 2008; Stone et al. 2000), few efforts have focused on aiding recall of substance use information. Hubbard (1992), however, has reported a series of experiments that used anchoring manipulations to improve

respondent recall, although these were not found to be very effective.

13.5.9 Response Editing

After respondents have interpreted a survey question and retrieved the information necessary to form an answer, they must decide whether that answer is to be accurately shared with the researcher. Given the illicit, and sometimes stigmatizing nature of substance use behaviors (Fortney et al. 2004), some respondents will make conscious decisions to underreport, or deny altogether, any such behavior (Hunt et al. 2015; Pernanen 1974). That survey respondents will sometimes attempt to present themselves in a favorable, albeit not completely accurate, light during survey interviews is well understood and is commonly referred to as social desirability bias. Concerns about the potential effects of social desirability bias have been the subject of considerable research in the survey methodology literature (Bradburn and Sudman 1979; Crowne and Marlowe 1964; Paulhus 1991; Tourangeau and Yan 2007). In general, respondents are known to over-report socially desirable behaviors, such as voting (Traugott and Kattosh 1979) and exercise (Adams et al. 2005), while underreporting socially undesirable behaviors, such as drug and alcohol use (Fendrich et al. 2004). Bradburn and Sudman (1979) have explored and documented the sensitive nature of substance use questions by asking a national sample of respondents in the U.S. how uneasy discussing various potentially sensitive topics would make them feel. They found that 42% reported that they believed most respondents would be "very uneasy" discussing their use of marijuana, and that 31 and 29%, respectively, would also be uneasy discussing stimulant and depressant use, and intoxication. Only 10% indicated they believed most people would be uneasy discussing drinking in general. It should be recognized that this survey was conducted more than 30 years ago, and it is uncertain to what degree these topics would elicit similar feelings of discomfort today.

Some respondents will feel uneasy discussing their substance use experiences. This may be for several reasons, including the need to avoid the social threat of admitting to illegal activities, and the feelings of shame and embarrassment associated with violating social norms (Harrel 1997; Krumpal 2013). Reporting illicit substance use may also be viewed by some respondents as a sign of weakness and, hence, something not to disclose (Robbins and Clayton 1989). These points are consistent with research findings that indicate that substance use underreporting increases with the perceived stigma of the substance being discussed (O'Malley et al. 1983; Mieczkowski 1989; Hser 1997). Some respondents may also elect not to admit to substance use behaviors in order to avoid potential legal sanctions, out of fear that a breach of confidentiality might risk their employment or reputation, and/or because they believe such information is highly personal and not to be shared. Some research suggests that questions about current use of illicit substances are more likely to produce underestimates when confidentiality is less certain, compared to questions concerned with past use (Luetgert and Armstrong 1973). Experimental studies that have compared substance use reporting patterns when provided with assurances of anonymity versus confidentiality have generally found few differences across conditions (Malvin and Moskowitz 1983; Moore and Ames 2002; O'Malley et al. 2000).

Measures of willingness to provide socially desirable answers have been found to be associated with substance use reporting such that likelihood of providing socially desirable responses in general is associated with less likelihood of reporting alcohol and/or drug use behavior (Bradburn and Sudman 1979; Pleck et al. 1996; Watten 1996; Welte and Russell 1993). These findings have been interpreted alternatively as evidence that underreporting of substance use is a consequence of respondent attempts to conceal illicit behavior, or as evidence that persons who engage in socially desirable behaviors in general also report, accurately, that they do not engage in substance use behaviors. Although this question remains

unresolved, we note that other research has demonstrated the absence of an association between one measure of social desirability, the Crowne–Marlowe scale (Crowne and Marlowe 1964) and a measure of cocaine use underreporting that was based on comparisons of self-reports with biological assays (Johnson et al. 2012).

As with question interpretation, respondent culture may also be associated with the accuracy of self-reports of substance use behaviors. A literature review of 36 published studies conducted in the U.S. found consistent evidence of lower reliability and validity rates of substance use reporting among racial and ethnic minority populations (Johnson and Bowman 2003). More recent studies have reported similar findings (Fendrich and Johnson 2005; Fendrich et al. 2008; Ledgerwood et al. 2008). The specific source of these differences, however, is not clearly understood. Models that have been proposed suggest that greater reporting errors among minority groups may be a consequence of differential group educational achievement and question comprehension, greater minority concerns with privacy, discrimination and risk of prosecution, and/or stronger effects of social desirability pressures on minority groups to report behaviors that conform to majority cultural values. Internationally, cultural differences in normative patterns of alcohol consumption and other substance use may also influence degree of response editing. In nations where wine is considered part of a meal, rather than mood-altering substance, underreporting might be expected to be much less of a concern.

An important limitation of much of the research reviewed here is the common, if unproven, assumption that greater self-reports of substance use behaviors are more valid (Del Boca and Darkes 2003; Gmel et al. 2014; Miller 1997). Over-reporting also needs to be recognized as a potential measurement concern (Harrison 2001; Miller 1997). There have been cases of respondents providing daily alcohol use reports that are physically impossible (Pernanen 1974). In surveys of adolescents, there is also a wide-spread belief that some respondents over-report their alcohol and other drug use, possibly to impress peers and improve one's social status, or as part of a general desire for attention (Barnea et al. 1987; Brener et al. 2003; Fendrich 2005; Midanik 1982; Percy et al. 2005; Swadi 1990). Gfroerer et al. (2012) suggest that such over-reporting of substance use might be more likely to happen during school-based surveys, typically conducted in classroom settings, where peers may be more likely to be aware of respondent answers. It is also possible that some respondents may in some situations elect to present themselves in a highly negative manner, perhaps for personal amusement or to obtain treatment services (de Lint 1981; Harris et al. 2008; Richter and Johnson 2001; Winters et al. 1991). To identify such over-reporters, several investigators have asked respondents about their use of substances that do not exist (Poulin et al. 1993). These studies have uniformly found very low self-reported rates of use of fictitious substances. Petzel et al. (1973), for example, reported that 4% of a high school student sample reported use of the non-existing drug "bindro." These researchers also found that those who indicated the use of a non-existent drug additionally reported more use of all other drugs included in their survey, compared to those who indicated, correctly, that they did not use "bindro." Others have reported similar findings when asking survey respondents about the use of non-existent substances (Barnea et al. 1987; Farrell et al. 1991; Single et al. 1975; Whitehead and Smart 1972). An alternative interpretation, of course, is that heavy drug users just assume, incorrectly, that they have used all available substances at one time or another and are hence trying to respond as accurately as possible.

Some have questioned whether or not it is correct to assume that all substance users will hesitate to accurately report their patterns of substance use. For example, Wish et al. (1997) have suggested that heavy substance users may be less concerned about social norms and consequences of reporting such information. Interviews with persons receiving treatment, though, have found little interest in publicly discussing their patterns of substance use (Willis 1997). Debates

regarding the accuracy of substance use reporting have led to numerous attempts to validate or corroborate survey responses. Several panel surveys, for example, have demonstrated considerable stability in respondent reporting of substance use over time (Bachman et al. 1981; Merline et al. 2008; Osgood et al. 1988). Research has also investigated the recanting of drug and alcohol use, which is defined as the tendency of some panel survey respondents to claim no lifetime experience with a given substance, when they have previously reported having used it (Fendrich 2005). Recanting has been identified in responses to both alcohol (Caldwell et al. 2006) and drug use questions (Fendrich and Rosenbaum 2003; Fendrich and Vaughn 1994; Fendrich and Mackesy-Amiti 2000; Johnston and O'Malley 1997; Percy et al. 2005; Shillington and Clapp 2000; Shillington et al. 2011a, b, c, d; Siddiqui et al. 1999).

Depending on the age group being surveyed (adults vs. adolescents), recanting may represent deliberate efforts to deny previously reported activity, exaggerations regarding behaviors that never actually took place, poor comprehension of survey questions during at least one wave of interviews, poor recall of information, or simple carelessness when answering (Fendrich 2005; Fendrich and Mackesy-Amiti 2000). Findings reported by Martino et al. (2009) suggest that recanting is a consequence both of deliberate misreporting and of errors in understanding of survey questions. In surveys of adolescents, one possible explanation for recanting is that younger and less-mature respondents may be more likely to exaggerate substance use during surveys conducted in classroom settings in which peers might be aware of one another's answers, and that they may then provide more accurate answers during subsequent survey waves as they subsequently become more mature (Fendrich and Rosenbaum 2003). Longitudinal follow-ups with Monitoring the Future survey respondents have found that recanting is greater among adults with occupations that might be expected to strongly sanction the use of illicit substances, such as those associated with the military and law enforcement (Johnston and O'Malley 1997). Percy et al. (2005) have also documented increased recanting among adolescents who had received drug education during the study period, suggesting a potentially biasing effect of education on self-reports. Higher recanting among low-level substance users has also been reported (Fendrich and Vaughn 1994; Percy et al. 2005).

Research has sought to validate self-reported substance use behavior by comparing those reports to toxicological findings from biospecimens collected at the time that interviews are conducted. One of the earliest studies conducted with a community sample (in Chicago) by Fendrich et al. (1999) found that recent cocaine and heroin use estimates obtained from hair testing were considerably higher than were self-reports obtained from the same respondents. A follow-up community survey found that higher rates of cocaine and heroin were obtained from drug assays of hair, saliva, and urine samples, compared to self-reports from respondents to a community survey (Fendrich et al. 2004). A higher estimate of marijuana use, though, was derived from self-reports, compared to drug test assays, a finding that was interpreted as evidence of the limitations of hair testing for the detection of marijuana use. Similar findings of underreporting of cocaine and heroin have also been obtained from general population surveys conducted in Puerto Rico by Colón et al. (2001, 2002, 2010), and of men who have sex with men in Chicago (Fendrich et al. 2008). Another study conducted as part of the NSDUH investigated agreement between self-reported use of marijuana and cocaine and urine tests, concluded that "most youths aged 12 to 17 and young adults aged 18 to 25 reported their recent drug use accurately" (Harrison et al. 2007). Ledgerwood et al. (2008) also examined the association between hair testing and self-reported illicit drug use in a community sample, concluding agreement between tests and self-reports to be substantial for marijuana and cocaine, moderate for opiates, and fair for methamphetamines. Other research has employed urinalysis (Morrall et al. 2000) and hair assays (Tassiopoulos et al. 2006) to document drug use frequency underreporting among drug users receiving treatment. While

providing valuable insights, it is important to acknowledge that each of these sources of confirmatory biological information is also imperfect measures of substance use, suffering from a variety of limitations, including imprecise and variable detection windows, vulnerability to contamination, and individual and race/ethnic group variability in rates of chemical absorption and retention (Wolff et al. 1999; DeLauder 2004).[6]

Confirmatory reports from social network members are another approach to validating self-reports of substance use obtained from respondents. This strategy has found good, but far from perfect levels of corroboration (Barnea et al. 1987; Leonard et al. 1983; Satyanarayana et al. 2010; Single et al. 1975; Sobell et al. 1997). Parents and children have also been asked to corroborate one another's reports of alcohol use. A study of African American parents and teenagers in Michigan concluded that both cocaine and opiate use by teenagers were underreported when self-reports were compared with confirmatory data provided by hair assays (Delaney-Black et al. 2010. In a Dutch study, Engels et al. (2007) also found that both children and parents underestimate one another's alcohol consumption, and that underestimation of adolescent alcohol consumption by parents was related to lack of knowledge and control of their children's activities. An important caveat to note when employing this type of approach is that proxy and self-reports generally suffer from the same sources of error (Del Boca and Noll 2000). Interestingly, perceptions of untrustworthiness by others have also been found to be associated with drug use recanting among adolescents in a study reported by Weinfurt and Bush (1996).

Another strategy for evaluating aggregate reporting of alcohol use is by way of comparison between alcohol sales and tax information. A number of studies have employed this approach and have consistently found evidence suggestive that survey self-reports underestimate total alcohol consumption (Kerr and Greenfield 2007; Nelson et al. 2010; Rehm 1998; Smith et al. 1990). State-level estimates (in the U.S.) from self-reports, however, correlate fairly strongly with the estimates from sales/tax data, suggesting sensitivity to variations in substance use behavior (Nelson et al. 2010). One study that compared self-reports of alcohol purchases, rather than self-reported alcohol consumption, found closer agreement between total estimates developed from those self-reports, in comparison to total retail alcohol sales in Sweden (Ramstedt 2010). This study also found considerable variability by alcohol type, with wine sales far more accurately reported than beer and spirits, suggesting the possibility that social desirability concerns may be at least partially responsible, given that wine is likely viewed as a more socially desirable alcoholic beverage in many social contexts. Reporting of wine consumption was also found to be more complete in a Canadian Study (Sims 1969).

Several strategies have been designed to address social desirability concerns in surveys. One such approach, intended to provide respondents with greater privacy when speaking with interviewers about highly sensitive questions, such as substance use behavior, is the randomized response technique, which was first proposed by Warner (1965). Several studies have documented the usefulness of this procedure among both students and adults. Goodstadt and Gruson (1975) found higher drug use reporting for five of six substances among high school students in Ontario. Weissman et al. (1986) compared substance use self-reports obtained with and without the use of the randomized response technique during telephone interviews conducted as part of a general household survey in New York City and also found increased reporting for three of four substances when using the randomized response technique. An important drawback noted, though, was that only 52% of those randomly assigned to respond using this technique actually agreed to do so. In contrast, McAuliffe et al. (1991) reported no differences in reports of illicit drug use among those responding via the randomized response technique, compared to those answering direct questions.

[6]Chapter 14 (this volume) by Fendrich et al. explores the use of biospecimens in greater detail.

Limitations of this technique include the challenge of correctly administering it in practice, as well as its ability to provide aggregate estimates only (Brener et al. 2003).

Another approach that has been employed to induce more accurate reporting of substance use behavior is the bogus pipeline. This involves the ethically questionable practice of leading respondents to believe their survey responses will be validated using some alternative means, when in fact the investigator has no intention of doing so. This implied threat of validating respondent answers is used to exert pressure on respondents to answer more truthfully. In general, the bogus pipeline procedure has failed to obtain higher estimates of substance use behavior, at least among adolescents (Campanelli et al. 1987; Murray and Perry 1987; Werch et al. 1987). A meta-analysis has confirmed the non-efficacy of the bogus pipeline procedure for improved reporting of alcohol consumption and marijuana use (Aguinis et al. 1995). One subsequent study, by Tourangeau et al. (1997), however, has demonstrated the effectiveness of the bogus pipeline technique for increasing respondent reporting of sensitive behaviors, including alcohol and illicit drug use. In addition, a special population study has suggested that the bogus pipeline procedure may be successful in improving self-reports under certain conditions. Lowe et al. (1986) found that, among pregnant women, those randomly assigned to a bogus pipeline condition were nearly twice as likely to report alcohol consumption when completing a self-administered questionnaire.

In addition, when considering respondent-related reporting errors, it is highly likely that multiple sources of respondent-related reporting errors are operating simultaneously. For example, Johnson and Fendrich (2005) demonstrated, using latent measures of cognitive processing difficulties constructed using debriefing probes, that social desirability concerns were predictive of discordant drug use reporting and drug use underreporting, while memory difficulties were predictive of drug use over-reporting.

13.5.10 Interviewer Effects

Survey interviewers can introduce errors by misreading questions, failing to probe answers correctly, not following other elements of standardized survey protocols, and by deliberate falsification of survey interviews (Johnson et al. 2001; Turner et al. 2002). Interviewer affiliation with governmental agencies may also influence respondent willingness to report substance use behaviors (Grucza et al. 2007). Interestingly, and somewhat counter-intuitively, interviewers with no prior project-related experience have been found to generate higher levels of marijuana and cocaine reporting in national substance use surveys (Hughes et al. 2002; Turner et al. 1992). One possibility is that more experienced interviewers may obtain fewer self-reports of substance use behaviors because they appear to be less likely to read questions exactly as worded, compared to new interviewers (Gfroerer et al. 2002). Research by Chromy et al. (2005) also finds that more experienced interviewers achieve higher response rates, in addition to eliciting fewer reports of substance use behaviors, suggesting they may be more successful in gaining cooperation from non-substance users who might find a survey on this topic to be less personally salient or interesting, although they do not believe this fully accounts for the observed differences, which remain unaccounted for.

Social distance is another possible mechanism that may account for some interviewer effects. It is possible that the social distance between respondents and interviewers may influence the willingness of some respondents to report sensitive behaviors such as substance use. Johnson et al. (2000) found that adult respondents in a telephone survey concerned with substance use treatment needs were more likely to report recent and lifetime drug use when respondent–interviewer dyads were characterized as having relatively little social distance. In that study, social distance was measured using a simple count of the number of shared demographic identities (same gender, same race/ethnicity, similar age,

similar educational attainment). Johnson et al. (1997) also explored the effects of social distance between race/ethnic groups in a study which probed respondents regarding how comfortable or uncomfortable they would feel when interviewed about their alcohol consumption patterns by interviewers from the same culture and from other cultural groups. When asked how they would feel if interviewed by an interviewer with the same background, large majorities of African American (88.8%), Mexican American (74.7%), Puerto Rican (85.9%), and non-Hispanic white (92.9%) respondents indicated they would feel comfortable. However, when asked how they would feel if the interviewer asking about their alcohol use was from another cultural group, the proportions indicating they would continue to feel comfortable decreased to 60.0% in African Americans and Mexican Americans, and 69.4% in Puerto Ricans. Among non-Hispanic whites, though, the proportion indicating they would continue to be comfortable remained very high (89.3%), suggesting group differences in reactions to interviewers of similar vs. different race/ethnic backgrounds.

The effects on substance use reporting of similarities and differences in various demographic characteristics between interviewers and respondents have also been examined in other research. In studies conducted in Iowa a number of years ago, female respondents were more likely to report alcohol consumption to male interviewers, and conversely, male respondents were more likely to report alcohol use to female interviewers (Mulford and Miller 1959). Johnson and Parsons (1994) found that homeless respondents were more likely to report drug use to male interviewers, a finding that they linked to a "likely user" hypothesis that suggests that male interviewers were more likely to elicit positive substance use reports because their gender is perceived as being more likely to be substance users themselves, and thus more tolerant of substance use by others. In contrast, a study conducted by Darrow et al. (1986) reported that gay males were more likely to report drug use to female interviewers, who were viewed as having greater empathy and sympathy for deviant behavior, compared to male interviewers. In a survey conducted in the Netherlands, higher rates of alcohol use were reported by Turkish and Moroccan respondents to Dutch interviewers, compared to interviewers who were ethnically matched (Dotinga et al. 2005). These researchers also hypothesized that minority respondents may have either exaggerated their alcohol consumption to comply with the perceived norms of the person interviewing them, or underreported, or denied altogether, the use of alcohol when interviewed by interviewers from an Islamic background who would have been perceived as having a far less permissive opinion of alcohol use. This limited evidence does not suggest a clear pattern of effects of any interviewer characteristics on respondent self-reports of substance use behaviors, although it does seem likely that interviewer characteristics do matter in many situations.

The familiarity of interviewers and respondents with one another may also influence the quality of self-reported substance use behaviors. For example, Mensch and Kandel (1988), in the panel survey of the National Longitudinal Survey of Youth, found that marijuana use reporting was lower among respondents who had been interviewed more times previously by the same interviewer, suggesting that interviewer familiarity cued respondents regarding social desirability expectations, which depressed their drug use reporting. It would again appear that, somewhat ironically, the use of experienced survey interviewers, a practice that would typically be considered an important strength, would appear in some circumstances to be a factor contributing to lower quality data, at least when interviewers are serially assigned to the same subsets of respondents.

13.5.11 Social Context

Aspects of the social and physical environment within which survey data are collected may also influence the quality of the information collected. An important aspect of the social environment during survey interviews that has received

attention is the absence or presence of other individuals during the interview, as this is believed to influence the social desirability demands or pressures that respondents may perceive. Overall, the presence of others during survey interviews is known to be associated with lower reporting of sensitive behaviors, including substance use. Wilson (1981) noted that, when interviews were conducted in the presence of another person, average weekly alcohol consumption was lower, compared to interviews conducted in private. Similar findings were reported by Edwards et al. (1998), but only among males. Several studies of adolescent reporting of alcohol and drug use also found that the presence of a parent during a household interview, perhaps not surprisingly, reduces respondent willingness to report such behaviors (Aquilino 1997; Aquilino et al. 2000; Gfroerer 1985; Hoyt and Chaloupka 1994; Schutz and Chilcoat 1994). Hoyt and Chaloupka (1994) reported that the presence of friends during an interview increased substance use reporting, and Aquilino et al. (2000) reported that the presence of a spouse or significant other had no effect on reports of alcohol and drug use. It is important to recognize, though, some potential confounding, as those most likely to have another person present during an interview are those who are married, and those who have children, and these variables are also commonly associated with less substance use behavior.

Social desirability pressures and self-report quality may also be influenced by the physical context within which interviews take place. Much of this evidence comes from comparisons of adolescent survey responses when the surveys are completed at home vs. in a school setting. In school settings, parental monitoring is likely to be perceived as less of a concern and confidentiality assurances likely to be more credible. Findings support this hypothesis, as Brener et al. (2006) and others (Gfroerer et al. 1997b, 2012; Kann et al. 2002; Rootman and Smart 1985) have reported that adolescents will underreport substance use during household surveys, relative to school-based surveys. However, Needle and colleagues (1983) and Zanes and Matsoukas (1979) did not find differences in the reports obtained from students in school- vs. home-based settings.

13.6 Processing Errors

Once data collection is complete, the construction of a final survey data set requires the implementation of numerous editing, screening, imputation, and weighting processes. The integrity of these processes is especially critical in substance use surveys, as they often involve assumptions regarding the substance use behaviors and reporting intentions of respondents. Research by Fendrich and Johnson (2001) has documented important differences in the assumptions underlying editing processes that are made across national surveys of substance use in the U.S. that can substantially influence the prevalence estimates generated by each. Researchers employ several techniques to screen completed substance use questionnaires for inclusion in final data files. For example, Farrell and colleagues (1991) have examined the effects of excluding respondents (1) who provided a large number of inconsistent answers, as well as those (2) who reported use of a fictitious substance, from final data sets. They concluded that effects of excluding these responses on prevalence estimates were minimal, although they cautioned that exclusionary criteria need to be used carefully in order to avoid producing non-representative results.

Data imputation processes can also have an important impact on survey estimates. For example, a past report by the U.S. General Accounting Office (1993) identified imputation problems in the National Household Survey on Drug Abuse in which the estimated number of past year heroin users in the U.S. ranged dramatically from 232,000 to 701,000, as a consequence of whether or not missing data imputation procedures were used. The same report also indicated that the sample weights used to construct subgroup estimates of the total number of illicit drug users were, in some instances, based on extremely small number of individuals in

some weighting cells who reported current drug use. In one case from the 1991 NHSDA, a single 79-year-old woman was projected to represent approximately 142,000 persons believed to have used heroin during the previous year. In such instances, a single erroneous data entry could have major effects on overall survey estimates.

Not properly accounting for a survey's sample design during data analysis can also have major effects on empirical findings. These might include instances in which (1) sample weights fail to incorporate all sample design and/or nonresponse factors, (2) when variances are unadjusted for the clustering of respondents within sampled geographic areas, or (3) when the available sample weights are not correctly used. One unfortunate example of the failure to properly employ sample weights occurred a decade ago when a report concerned with illegal sales of alcohol to underage minors in the U.S. seriously overestimated the proportion of all alcohol sales that were being made to underage youth. Researchers were conducting a secondary analysis of a public release version of the 1998 NHSDA and failed to weight their data for the survey's stratified sample design, in which young persons, ages 12–20, were significantly over-sampled. Because only persons under the age of 21 purchase alcohol illegally in the U.S., their over-representation in the unweighted NHSDA data file led to an over-representation of illegal sales in those data. This was an error that could have been avoided through the use of the sample weights that were available in the public use data file used for those analyses. Those erroneous findings were quickly discovered after being made public (Lewin 2002).

13.7 Conclusions

Over several decades, considerable knowledge has been accumulated regarding sources of error in the survey assessment of substance use behaviors. Important gaps remains, however, and continued research will be necessary. Below, important unresolved questions that are relevant to each source of survey error are briefly considered. Regarding coverage errors, the challenge of constructing representative sample frames for both adolescents and adults continues to increase as electronic communications platforms further diversify. This is a general problem that afflicts all survey research efforts, but one that can be particularly problematic for substance use research given the associations between these behaviors and likelihood of being covered by many of the potential sources of sample frames. Identification of supplemental frames that might provide better coverage of heavy substance users and which could be employed, with appropriate weights, as supplements to more traditional sample frames when conducting population surveys need to be explored. Use of supplemental frames to better reach minority and underrepresented groups also needs to be examined.

When survey estimates are reported, sampling errors, in the form of standard errors or confidence intervals, are commonly included. Although reporting these errors is important to survey transparency, it is important to recognize that sampling errors make strong assumptions that are seldom met in practice. Most importantly, they assume the absence of all other sources of survey error. Given the unlikelihood of this assumption, merely reporting sampling errors can leave survey consumers with a false sense of the precision of survey estimates, as any sampling errors could be completely overwhelmed by measurement and/or nonresponse errors, for example, in practice. Understanding how sampling errors in substance use surveys may be influenced by other sources of survey error thus seems to be an important research question to be addressed in the future. Similarly, nonresponse errors are another permanent concern that substance use surveys will need to continually address. Of course, the degree to which nonresponse may bias survey findings will vary from topic to topic, question to question, and subpopulations of interest. Given the strong associations detected between substance use and nonresponse patterns, it appears that this error source is also particularly relevant for surveys on this topic. An important issue for additional research is the relative usefulness for substance

use surveys of the various nonresponse bias analytic strategies reviewed earlier in this paper. Similarly, research into the relative efficacy of various types of adjustments for nonresponse and other forms of error in substance use surveys would seem to be an important future research topic.

It is likely that the multiple sources of measurement errors reviewed earlier in this chapter pose the greatest threat to the accurate assessment of substance use behaviors. There are several practical questions that remain unresolved, however, such as the predictive power of social desirability measures, the reasons why experienced interviewers appear to obtain fewer reports of substance use behaviors, and the degree to which adolescents might actually over-report their use of alcohol and/or other drugs. Perhaps even more important, how these widely diverse sets of measurement errors interact with one another is poorly understood and remains largely unexamined. Evaluation of how various sources of measurement errors in substance use surveys interact together to influence survey estimates should be a priority for future research.

In terms of processing errors, surveys concerned with substance use would appear, on the surface, to be no more vulnerable than other types of survey research. Yet, the complexity of most substance use questionnaires, combined with greater item nonresponse rates in many instances, likely provide greater risks for processing errors that can be linked to complex editing rules and assumptions. A general rule of thumb is that the likelihood of experiencing processing errors is inversely associated with the amount of documentation provided with a survey, as careful documentation is an important indicator of quality research. Continued research into the veracity of data editing decision rules, particularly when handling missing data and/or inconsistent self-reports in substance use surveys, would certainly be welcomed.

Finally, it is strongly recommended that substance use researchers who plan to employ survey research methods recognize and report on their efforts to address each of these potential sources of survey-related error. Developing strategies to systematically and rigorously confront each source of error, and transparently sharing one's successes and failures, remains the best approach to minimizing the effects of each when using survey methods to investigate substance use patterns and behaviors.

Acknowledgements An earlier version of this chapter was published as Johnson (2014) Sources of error in substance use prevalence surveys, International Scholarly Research Notices, Article ID 923290.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, J. E., Cunningham, J., Fulton, J., et al. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology, 161,* 389–398.

Aguinis, H., Pierce, C. A., & Quigley, B. M. (1995). Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: A meta-analytic review. *Basic and Applied Social Psychology, 16,* 515–527.

Andersen, E. M., Diehr, P. H., & Luke, D. A. (2004). Public health surveillance of low-frequency populations. *Annual Review of Public Health, 25,* 25–52.

Aquilino, W. S. (1992). Telephone versus face-to-face interviewing for household drug use surveys. *International Journal of the Addictions, 27,* 71–91.

Aquilino, W. S. (1994). Interviewer mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly, 58,* 210–240.

Aquilino, W. S. (1997). Privacy effects on self-reported drug use: Interactions with survey mode and respondent characteristics. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No National Institute on Drug Abuse.

Aquilino, W. S., & LoSciuto, L. A. (1990). Effects of interview mode of self-reported drug use. *Public Opinion Quarterly, 54,* 362–395.

Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse, 35,* 845–846.

Ardila, A., Rosselli, M., & Strumwassler, S. (1991). Neuropsychological defects in chronic cocaine abuse. *International Journal of Neuroscience, 57,* 73–79.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. K. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol, 61,* 55–63.

Bachman, J. G., Johnson, L. D., & O'Malley, P. M. (1981). Smoking, drinking, and drug use among American high school students: Correlates and trends, 1975–1979. *American Journal of Public Health, 71,* 59–69.

Bachman, J. G., & O'Malley, P. M. (1981). When four months equal a year: Inconsistencies in students' reports of drug use. *Public Opinion Quarterly, 45,* 536–548.

Bailey, S. L., Flewelling, R. L., & Rachal, J. V. (1992). The characterization of inconsistencies in self-reports of alcohol and marijuana use in a longitudinal study of adolescents. *Journal of Studies on Alcohol, 53,* 636–647.

Barnea, A., Rahav, G., & Teichman, M. (1987). The reliability and consistency of self-reports on substance use in a longitudinal study. *British Journal of Addiction, 82,* 891–898.

Barrett, S. P., Gross, S. R., Garand, I., & Pihl, R. O. (2005). Patterns of simultaneous polysubstance use in Canadian rave attendees. *Substance Use and Misuse, 40,* 1525–1537.

Bauermeister, J. A., Zimmerman, M. A., Johns, M. M., Glowacki, P., Stoddard, S., & Volz, E. (2012). Innovative recruitment using online networks: Lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. *Journal of Studies on Alcohol and Drugs, 73,* 834–838.

Beard, C. M., Lane, A. W., O'Fallon, W. M., Riggs, B. L., & Melton, L. S. (1994). Comparison of respondents and nonrespondents in an osteoporosis study. *Annals of Epidemiology, 4,* 398–403.

Belli, R. F. (2008). *Calendar and time diary methods in life course research*. Los Angeles, CA: Sage Publications.

Bijl, R. V., van Zessen, G., Ravelli, A., de Rijk, C., & Langendoen, Y. (1998). The Netherlands Mental Health Survey and Incidence Study (NEMESIS): Objectives and design. *Social Psychiatry and Psychiatric Epidemiology, 33,* 581–586.

Bloomfield, K., Hope, A., & Kraus, L. (2013). Alcohol survey measures for Europe: A literature review. *Drugs: Education Prevention and Policy, 20,* 348–360.

Blumberg, S. J., & Luke, J. V. (2009). Reevaluating the need for concern regarding noncoverage bias in landline surveys. *American Journal of Public Health, 99,* 1806–1810.

Blumberg, S. J., & Luke, J. V. (2016). *Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2015*. Hyattsville, MD: National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf

Blumberg, S. J., Luke, J. V., & Cynamon, M. L. (2006). Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution. *American Journal of Public Health, 96,* 926–931.

Bolla, K. I., McCann, U. D., & Ricaurte, G. A. (1998). Memory impairment in abstinent MDMA ("Ecstasy") users. *Neurology, 51,* 1532–1537.

Bongers, I. M. B., & Oers, J. A. M. (1998). Mode effects on self-reported alcohol use and problem drinking: Mail questionnaires and personal interviewing compared. *Journal of Studies on Alcohol, 59,* 280–285.

Boniface, S., Kneale, J., & Shelton, N. (2014). Drinking pattern is more strongly associated with under-reporting of alcohol consumption than socio-demographic factors: Evidence from a mixed-methods study. *BMC Public Health*, 14.

Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. San Francisco, CA: Jossey-Bass Publishers.

Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33,* 436–457.

Brener, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., et al. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opinion Quarterly, 70,* 354–374.

Bucholz, K. K., Shayka, J. J., Marion, S. L., Lewis, C. E., Pribor, E. F., & Rubio, D. M. (1996). Is a history of alcohol problems or of psychiatric disorder associated with attrition at 11-year follow-up? *Annals of Epidemiology, 6,* 228–234.

Caetano, R., Ramisetty-Mikler, S., & McGrath, C. (2003). Characteristics of non-respondents in a US national longitudinal survey on drinking and intimate partner violence. *Addiction, 98,* 791–797.

Caldwell, T. M., Rogers, B., Power, C., Clark, C., & Stansfeld, S. A. (2006). Drinking histories of self-identified lifetime abstainers and occasional drinkers: Findings from the 1958 British Birth Cohort Study. *Alcohol and Alcoholism, 41,* 650–654.

Campanelli, P., Dielman, T. E., & Shope, J. T. (1987). Validity of adolescents' self-reports of alcohol use and misuse using a bogus pipeline procedure. *Adolescence, 22,* 7–22.

Caspar, R. A. (1992). Follow-up of non-respondents in 1990. In C. Turner, J. Lessler & J. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies*. DHHS Pub. No., ADM 92-1929. Rockville, MD: National Institute on Drug Abuse.

Center for Behavioral Health Statistics and Quality. (2015). *Behavioral health trends in the United States: Results from the 2014 national survey on drug use and health*. HHS Publication No. SMA 15-4927, NSDUH Series H-50. Retrieved from http://www.samhsa.gov/data/

Chavez, E. L., Edwards, R. W., & Oetting, E. R. (1989). Mexican American and White American school dropouts drug use, health status, and involvement in violence. *Public Health Reports, 104,* 594–604.

Cho, Y. I., Johnson, T. P., & Fendrich, M. (2001). Monthly variations in self-reports of alcohol consumption. *Journal of Studies on Alcohol, 62,* 268–272.

Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues*. DHHS Publication No. MA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Chromy, J. R., Eyerman, J., Odom, D., & McNeeley, A. E. (2005). Association between interviewer experience and substance use prevalence rates in NSDUH. In J. Kannet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey of Drug Use and Health*. DHHS Publication No. SMA 05-4044, Methodology Series M-5. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Chung, T., & Martin, C. S. (2005). What were they thinking? Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug and Alcohol Dependence, 80,* 191–200.

Clatts, M. C., Goldamt, L. A., & Yi, H. (2005). Club drug use among young men who have sex with men in NYC: A preliminary epidemiologic profile. *Substance Use and Misuse, 40,* 1317–1330.

Cohen, G., & Duffy, J. C. (2002). Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics, 18,* 13–23.

Colón, H. M., Pérez, C. M., Meléndez, M., Marrero, E., Ortiz, A. P., & Suárez, E. (2010). The validity of drug use responses in a household survey in Puerto Rico: Comparison of survey responses with urinalysis. *Addictive Behaviors, 35,* 667–672.

Colón, H. M., Robles, R. R., & Sahai, H. (2001). The validity of drug use responses in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *International Journal of Epidemiology, 30,* 1042–1049.

Colón, H. M., Robles, R. R., & Sahai, H. (2002). The validity of drug use reports among hard core drug users in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *Drug and Alcohol Dependence, 67,* 269–279.

Cottler, L. B., Zipp, J. F., Robbins, L. H., & Spitznagel, E. L. (1987). Difficult-to-recruit respondents and their effect on prevalence estimates in an epidemiologic survey. *American Journal of Epidemiology, 125,* 329–339.

Cowan, C. D. (2001). Coverage, sample design, and weighting in three federal surveys. *Journal of Drug Issues, 31,* 599–614.

Crawford, A. (1986). A comparison of participants and non-participants from a British general population survey of alcohol drinking practices. *Journal of the Market Research Society, 28,* 291–297.

Crawford, A. (1987). Bias in a survey of drinking habits. *Alcohol and Alcoholism, 22,* 167–179.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence New York*. NY: Wiley.

Cunradi, C. B., Moore, R., Killoran, M., & Ames, G. (2005). Survey nonresponse bias among young adults: The role of alcohol, tobacco, and drugs. *Substance Use and Misuse, 40,* 171–185.

Darrow, W. W., Jaffe, H. W., Thomas, P. A., Haverkos, H. W., Rogers, M. F., Guinan, M. E., et al. (1986). Sex of interviewer, place of interview, and responses of homosexual men to sensitive questions. *Archives of Sexual Behavior, 15,* 79–88.

Dawson, D. A. (1998). Volume of ethanol consumption: Effects of different approaches to measurement. *Journal of Studies on Alcohol, 59,* 191–197.

Dawson, D. A. (2003). Methodological issues in measuring alcohol use. *Alcohol Research & Health, 27,* 18–29.

Dawson, D. A., Goldstein, R. B., Pickering, R. P., & Grant, B. F. (2014). Nonresponse bias in survey estimates of alcohol consumption and its association with harm. *Journal of Studies on Alcohol and Drugs, 75,* 695–703.

De Graf, R., van Dorsselaer, S., Tuithof, M., & ten Have, M. (2013). Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands mental health survey and incidence study-2. *Comprehensive Psychiatry, 54,* 1131–1139.

De La Rosa, M. R., & Adrados, J.-L. R. (1993). *Drug abuse among minority youth: Advances in research and methodology*. Rockville, MD: NIDA Research Monograph 130, National Institute on Drug Abuse.

de Lint, J. (1981). "Words and deeds": Responses to Popham and Schmidt. *Journal of Studies on Alcohol, 42,* 359–361.

DeBlaere, C., Brewster, M. E., Sarkees, A., & Moradi, B. (2010). Conducting research with LGB people of color: Methodological challenges and strategies. *The Counseling Psychologist, 38,* 331–362.

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction, 98* (Supplement 2), 1–12.

Del Boca, F. K., & Noll, J. A. (2000). Truth or consequences: The validity of self-report data in health services research. *Addiction, 95*(Supplement 3), S347–S360.

Delaney-Black, V., Chiodo, L. M., Hannigan, J. H., Greenwald, M. K., Janisse, J., Patterson, G., et al. (2010). Just say "I don't": Lack of concordance between teen report and biological measures of drug use. *Pediatrics, 126,* 887–893.

DeLauder, S. F. (2004). Considering issues of racial bias in drug testing where hair is the matrix. *Transforming Anthropology, 11,* 54–59.

Delnevo, C. D., Gundersen, D. A., & Hagma, B. T. (2007). Declining estimated prevalence of alcohol

drinking and smoking among young adults nationally: Artifacts of sample undercoverage. *American Journal of Epidemiology, 167,* 15–19.

Dengler, R. (1996). Smoking and alcohol consumption in Trent, UK: An analysis of item non-response. *Journal of Epidemiology and Community Health, 50,* 687.

Devos-Comby, L., & Lange, J. E. (2008). "My drink is larger than yours"? A literature review of self-defined drink sizes and standard drinks. *Current Drug Abuse Reviews, 1,* 162–176.

Dotinga, A., van den Eijnden, R. J. J. M., Bosvekd, W., & Garretsen, H. F. L. (2005). The effect of data collection and ethnicity of interviewer on response rates and self-reported alcohol use among Turks and Moroccans in the Netherlands: An experimental study. *Alcohol and Alcoholism, 40,* 242–248.

Duffy, J. C., & Waterton, J. J. (1984). Under-reporting of alcohol consumption in sample surveys: The effect of computer interviewing in fieldwork. *British Journal of Addiction, 79,* 303–308.

Dunne, M. P., Martin, N. G., Bailey, J. M., Heath, A. C., Bucholz, K. K., Madden, P. A. F., et al. (1997). Participation bias in a sexuality survey: Psychological and behavioural characteristics of responders and non-responders. *International Journal of Epidemiology, 26,* 844–854.

Eaton, D. K., Brener, N. D., Kann, L., Denniston, M. M., McManus, T., Kyle, T. M., et al. (2010). Comparison of paper-and-pencil versus web administration of the Youth Risk Behavior Survey (YRBS): Risk behavior prevalence estimates. *Evaluation Review, 34,* 137–153.

Edwards, S. L., Slattery, M. L., & Ma, K.-N. (1998). Measurement errors stemming from nonrespondents present at in-person interviews. *Annals of Epidemiology, 8,* 272–277.

Ekholm, O. (2004). Influence of the recall period on self-reported alcohol intake. *European Journal of Clinical Nutrition, 58,* 60–62.

Ekholm, O., Strandberg-Larsen, K., Christensen, K., & Grønbæk, M. (2008). Comparison of assessment methods for self-reported alcohol consumption in health interview surveys. *European Journal of Clinical Nutrition, 62,* 286–291.

Ellickson, P., Bui, K., Bell, R., & McGuigan, K. (1998). Does early drug use increase the risk of dropping out of high school? *Journal of Drug Issues, 28,* 357–380.

Elliott, M. N., Finch, B. K., Klein, D., Ma, S., Do, D. P., Beckett, M. K., et al. (2008). Sample designs for measuring the health of small racial/ethnic subgroups. *Statistics in Medicine, 27,* 4016–4029.

Engels, R., Knibbe, R. A., & Drop, M. J. (1997). Inconsistencies in adolescents' self-reports of initiation of alcohol and tobacco use. *Addictive Behaviors, 22,* 613–623.

Engels, R. C. M. E., Van Der Vorst, H., Dekoić, M., & Meeus, W. (2007). Correspondence in collateral and self-reports on alcohol consumption: A within family analysis. *Addictive Behaviors, 32,* 1016–1030.

Farrell, A. D., Danish, S. J., & Howard, C. W. (1991). Evaluation of data screening methods in surveys of adolescents' drug use. *Psychological Assessment, 3,* 295–298.

Fendrich, M. (2005). The undeniable problem of recanting. *Addiction, 100,* 143–144.

Fendrich, M., & Johnson, T. P. (2001). Examining prevalence differences in three national surveys of youth: Impact of consent procedures, mode, and editing rules. *Journal of Drug Issues, 31,* 615–642.

Fendrich, M., & Johnson, T. P. (2005). Race/ethnicity differences in the validity of self-reported drug use: Results from a household survey. *Journal of Urban Health*, 82 Supplement 3, iii67–iii81.

Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. S., & Spiehler, V. (1999). Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests. *American Journal of Epidemiology, 149,* 955–962.

Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A., & Spiehler, V. (2004). The utility of drug testing in epidemiological research: Results from an ACASI general population survey. *Addiction, 99,* 197–208.

Fendrich, M., & Mackesy-Amiti, M. E. (2000). Decreased drug reporting in a cross-sectional student drug use survey. *Journal of Substance Abuse, 11,* 161–172.

Fendrich, M., Mackesy-Amiti, M., & Johnson, T. P. (2008). Validity of self-reported substance use in men who have sex with men: Comparisons with a general population sample. *Annals of Epidemiology, 18,* 752–759.

Fendrich, M., & Rosenbaum, D. P. (2003). Recanting of substance use reports in longitudinal prevention study. *Drug and Alcohol Dependence, 70,* 241–253.

Fendrich, M., & Vaughn, C. (1994). Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opinion Quarterly, 58,* 96–123.

Fernández, M. L., Bowen, G. S., Varga, L. M., Collazo, J. B., Hernandez, N., Perrino, T., et al. (2005). High rates of club drug use and risky sexual practices among Hispanic men who have sex with men in Miami, Florida. *Substance Use and Misuse, 40,* 1347–1362.

Fortney, J., Mukherjee, S., Curran, G., Fortney, S., Han, X., & Booth, B. M. (2004). Factors associated with perceived stigma for alcohol use and treatment among at-risk drinkers. *Journal of Behavioral Health Services & Research, 31,* 418–429.

Fowler, F. J., & Stringfellow, V. L. (2001). Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols. *Journal of Drug Issues, 31,* 643–664.

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin, 113,* 44–66.

Garcia, M., Fernandez, E., Schiaffino, A., Borrell, C., Marti, M., & Borras, J. M. (2005). Attrition in a population-based cohort eight years after baseline interview: The Conella Health Interview Survey Follow-up (CHIS.FU) study. *Annals of Epidemiology, 15,* 98–104.

Gardner, B., & Tang, V. (2014). Reflecting on non-reflective action: An exploratory think-aloud study of self-report habit measures. *British Journal of Health Psychology, 19,* 258–273.

Gfroerer, J. (1985). Underreporting of drug use by youths resulting from lack of privacy in household interviews. In B. Rouse, N. Kozel, & L. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity*. National Institute on Drug Abuse: Washington, D.C.

Gfroerer, J., Bose, J., Kroutil, L., Lopez, M., & Kann, L. (2012). Methodological considerations in estimating adolescent substance use. In *Proceedings from the 2012 Joint Statistical Meetings, Section on Survey Research Methods* (pp. 4127–4140).

Gfroerer, J., Eyerman, J., & Chromy, J. (2002). *Redesigning an ongoing national household survey: Methodological issues*. DHHS Pub. No. SMA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Gfroerer, J., & Hughes, A. (1991). The feasibility of collecting drug abuse data by telephone. *Public Health Reports, 106,* 284–293.

Gfroerer, J., & Hughes, A. (1992). Collecting data on illicit drug use by phone. In C. Turner, J. Lessler, & J. Gfroerer (Eds.) *Survey measurement of drug use: Methodological studies*. DHHS Pub. No., ADM 92-1929. Rockville, MD: National Institute on Drug Abuse.

Gfroerer, J., Lessler, J., & Parsley, T. (1997a). Studies of nonresponse and measurement error in the National Households Survey on Drug Abuse. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Gfroerer, J., Wright, D., & Kopstein, A. (1997b). Prevalence of youth substance use: The impact of methodological differences between two national surveys. *Drug and Alcohol Dependence, 47,* 19–30.

Gmel, G. (2000). The effect of mode of data collection and of non-response on reported alcohol consumption: A split-sample study in Switzerland. *Addiction, 95,* 123–134.

Gmel, G., & Daeppen, J.-B. (2007). Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: Implications for case-crossover designs. *Journal of Studies on Alcohol and Drugs, 68,* 303–310.

Gmel, G., Graham, K., Kuendig, H., & Kuntsche, S. (2006). Measuring alcohol consumption—Should the 'graduated frequency' approach become the norm in survey research? *Addiction, 101,* 16–30.

Gmel, G., & Rehm, J. (2004). Measuring alcohol consumption. *Contemporary Drug Problems, 31,* 467–540.

Gmel, G., Studer, J., Deline, S., Baggio, S., N'Goran, A., Mohler-Kuo, M., et al. (2014). More is not always better—Comparison of three instruments measuring volume of drinking in a sample of young me and their association with consequences. *Journal of Studies on Alcohol and Drugs, 75,* 880–888.

Goldberg, M., Chastang, J. E., Zins, M., Niedhammer, I., & Leclerc, A. (2006). Health problems were the strongest predictors of attrition during follow-up of the GAZEL cohort. *Journal of Clinical Epidemiology, 59,* 1213–1221.

Goodstadt, M. S., & Grusin, V. (1975). The randomized response technique: A test on drug use. *Journal of the American Statistical Association, 70,* 814–818.

Gorman, E., Leyland, A. H., McCartney, G., White, I. R., Katikireddi, S. V., Rutherford, L., et al. (2014). Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *American Journal of Epidemiology, 180,* 941–948.

Grant, B. F., & Dawson, D. A. (1998). Age of onset of drug use and its association with DSMIV drug abuse and dependence: Results from the national longitudinal alcohol epidemiologic survey. *Journal of Substance Abuse, 10,* 163–173.

Grant, B. F., Harford, T. C., Dawson, D. A., Chou, P. S., & Pickering, R. P. (1995). The alcohol use disorder and associated disabilities interview schedule (AUDADIS): Reliability of alcohol and drug modules in a general population sample. *Drug and Alcohol Dependence, 39,* 37–44.

Greenfield, T. K., & Kerr, W. C. (2008). Alcohol measurement methodology in epidemiology: Recent advances and opportunities. *Addiction, 103,* 1082–1099.

Greenfield, T. K., Midanik, L. T., & Rogers, J. D. (2000). Effects of telephone versus face-to-face interview modes on reports of alcohol consumption. *Addiction, 95,* 277–284.

Gribble, J. N., Miller, H. G., Cooley, P. C., Catania, J. A., Pollack, L., & Turner, C. F. (2000). The impact of T-ACASI interviewing on reported drug use among men who have sex with men. *Substance Use and Misuse, 36,* 869–890.

Groves, R. M. (1989). *Survey errors and survey costs*. New York, NY: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household surveys*. New York, NY: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (2002). *Survey nonresponse*. New York, NY: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72,* 167–189.

Grucza, R. A., Abbacchi, A. M., Przybeck, T. R., & Gfroerer, J. C. (2007). Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. *Addiction, 102,* 623–629.

Guttmacher, S., Weitzman, B. C., Kapadeia, F., & Weinberg, S. L. (2002). Classroom-based surveys of adolescent risk-taking behaviors: Reducing the bias of

absenteeism. *American Journal of Public Health, 92,* 235–237.

Halkitis, P. N., Fischgrund, B. N., & Parsons, J. T. (2005). Explanations for methamphetamine use among gay and bisexual men in New York City. *Substance Use and Misuse, 40,* 1331–1345.

Hall, W. (2014). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction, 110,* 19–35.

Hansen, W. B., Collins, L. M., Malotte, C. K., Johnson, C. A., & Fielding, J. E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine, 8,* 261–275.

Harrel, M. (1997). The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Harris, K. M., Griffin, B. A., McCaffrey, D. F., & Morral, A. R. (2008). Inconsistencies in self-reported drug use by adolescents in substance abuse treatment: Implications for outcome and performance measurements. *Journal of Substance Abuse Treatment, 34,* 347–355.

Harrison, L. D. (2001). Understanding the differences in youth drug prevalence rates produced by the MTF, NHSDA, and YRBS studies. *Journal of Drug Issues, 31,* 665–694.

Harrison, L. D., Martin, S. S., Enev, T., & Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population*. DHHS Pub. No. SMA 07-4249, Methodology Series M-7. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Hasin, D., & Carpenter, K. M. (1998). Difficulties with questions on usual drinking and the measurement of alcohol consumption. *Alcoholism, Clinical and Experimental Research, 22,* 580–584.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44,* 174–199.

Heckathorn, D. D. (2002). Respondent driven sampling II. Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems, 49,* 11–34.

Hill, A., Roberts, J., Ewings, P., & Gunnell, D. (1997). Non-response bias in a lifestyle survey. *Journal of Public Health Medicine, 19,* 203–207.

Hilton, M. E. (1989). A comparison of a prospective diary and two summary recall techniques for recording alcohol consumption. *British Journal of Addiction, 84,* 1085–1092.

Hines, D. A., Douglas, E. M., & Mahmood, S. (2010). The effects of survey administration on disclosure rates to sensitive items among men: A comparison of an internet panel sample with a RDD telephone sample. *Computers in Human Behavior, 26,* 1327–1335.

Hochhauser, M. (1979). Bias in drug abuse survey research. *International Journal of the Addictions, 14,* 675–687.

Hoeymans, N., Feskens, E. J. M., van den Bos, G. A. M., & Kromhout, D. (1998). Non-response bias in a study of cardiovascular diseases, functional status and self-rated health among elderly men. *Age and Ageing, 27,* 35–40.

Horwood, L., Fergusson, D., Havatbakksh, M., Najman, J., Coffey, C., Patton, G. C., et al. (2010). Cannabis use and educational achievement: Findings from three Australasian cohort studies. *Drug and Alcohol Dependence, 110,* 247–253.

Hoyt, G. M., & Chaloupka, F. J. (1994). Effect of survey conditions on self-reported substance use. *Contemporary Economic Policy, 12,* 109–121.

Hser, Y. (1997). Self-reported drug use: Results of selected empirical investigations of validity. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Hu, S. S., Balluz, L., Battaglia, M. P., & Frankel, M. R. (2011). Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. *American Journal of Epidemiology, 173,* 703–711.

Hubbard, M. (1992). Laboratory experiments testing new questioning strategies. In C. F. Turner, J. T. Lessler, & J.C. Gfroerer (Eds.), *Survey measurement of drug use*. Rockville, MD: National Institute on Drug Abuse, Rockville.

Hubbard, M., Pantula, J., & Lessler, J. (1992). Effects of decomposition of complex concepts. In C. F. Turner, J. T. Lessler & J. C. Gfroerer (Eds.), *Survey measurement of drug use.* Rockville, MD: National Institute on Drug Abuse.

Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2002). Impact of interviewer experience on respondent reports of substance use. In J. Gfroerer, J. Eyeman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues*. DHHS Publication No. SMA 03-3768, pp. 161-184. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Humphrey, J. A., & Friedman, J. (1986). The onset of drinking and intoxification among university students. *Journal of Studies on Alcohol, 47,* 455–458.

Hunt, D. E., Kling, R., Almozlino, Y., Jalbert, S., Chapman, M. T., & Rhodes, W. (2015). Telling the truth about drug use: How much does it matter. *Substance Abuse, 45,* 314–329.

Iversen, L., & Klausen, H. (1986). Alcohol consumption among laid-off workers before and after closure of a Danish ship-yard: A 2-year follow-up study. *Social Science and Medicine, 22,* 107–109.

Jabine, T., Straf, M., Tanur, J., & Tourangeau, R. (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, D.C.: National Academy Press.

Johnson, T. P. (2014). Sources of error in substance use prevalence surveys. *International scholarly*

research notices, Article ID 923290. Retrieved from: http://dx.doi.org/10.1155/2014/923290

Johnson, T. P., & Bowman, P. J. (2003). Cross-cultural sources of measurement error in substance use surveys. *Substance Use and Misuse, 38,* 1447–1490.

Johnson, T., & Fendrich, M. (2005). Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of Epidemiology, 15,* 381–389.

Johnson, T. P., Fendrich, M., & Mackesy-Amiti, M. E. (2012). An evaluation of the validity of the Crowne-Marlowe need for approval scale. *Quality & Quantity, 46,* 1883–1896.

Johnson, T. P., Fendrich, M., Shaligram, C., Garcy, A., & Gillespie, S. (2000). An evaluation of the effects of interviewer characteristics in an RDD telephone survey of drug use. *Journal of Drug Issues, 30,* 77–102.

Johnson, T. P., Hougland, J., & Clayton, R. (1989). Obtaining reports of sensitive behaviors: A comparison of substance use reports from telephone and face-to-face interviews. *Social Science Quarterly, 70,* 174–183.

Johnson, T. P., & Mott, J. A. (2001). The reliability of self-reported age of onset of tobacco, alcohol, and illicit drug use. *Addiction, 96,* 1187–1198.

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L., et al. (1997). Social cognition and responses to survey questions among culturally diverse populations. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality*. New York, NY: Wiley.

Johnson, T. P., & Owens, L. (2004). Survey response rate reporting in the professional literature. In *2003 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Johnson, T. P., Parker, V., & Clements, C. (2001). Detection and prevention of data falsification in survey research. *Survey Research, 32,* 1–2.

Johnson, T. P., & Parsons, J. A. (1994). Interviewer effects on self-reported substance use among homeless persons. *Addictive Behaviors, 19,* 83–93.

Johnston, L. D., & O'Malley, P. M. (1997). The recanting or earlier reported drug use by young adults. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Jousilahti, P., Salomaa, V., Kuulasmaa, K., Niemelä, M., & Vartiainen, E. (2005). Total and cause specific mortality among participants and non-participants of population based health surveys: A comprehensive follow up of 54 372 Finnish men and women. *Journal of Epidemiology and Community Health, 59,* 310–315.

Kalsbeek, W. D. (2003). Sampling minority groups in health surveys. *Statistics in Medicine, 22,* 1527–1549.

Kandel, D. B. (1975). Reaching the hard-to-reach: Illicit drug use among high school absentees. *Addictive Diseases, 1,* 465–480.

Kann, L., Brener, N. D., Warren, C. W., Collins, J. L., & Giovino, G. A. (2002). An assessment of the effect of data collection setting on the prevalence of health-risk behaviors among adolescents. *Journal of Adolescent Health, 31,* 327–335.

Kaplan, C. D., Korf, D., & Sterk, C. (1987). Temporal and social contexts of heroin-using populations: An illustration of the snowball sampling technique. *The Journal of Nervous and Mental Disease, 175,* 566–574.

Kassira, E. N., Bauserman, R. L., Tomoyasu, N., Caldeira, E., Swetz, A., & Solomon, L. (2001). HIV and AIDS surveillance among inmates in Maryland prisons. *Journal of Urban Health, 78,* 256–263.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a telephone survey. *Public Opinion Quarterly, 64,* 125–148.

Kerr, W. C., & Greenfield, T. K. (2007). Distribution of alcohol consumption and expenditures and the impact of improved measurement on coverage of alcohol sales in the 2000 National Alcohol Survey. *Alcohol: Clinical and Experimental Research*, *31*, 1714–1722.

Kerr, W. C., & Stockwell, T. (2012). Understanding standard drinks and drinking guidelines. *Drug and Alcohol Review, 21,* 200–205.

Khadjesari, Z., Murray, E., Kalaitzaki, E., White, I. R., McCambridge, J., Gadfrey, C., et al. (2009). Test-retest reliability of an online measure of past week alcohol consumption (The TOT-AL), and comparison with face-to-face interview. *Addictive Behaviors, 34,* 337–342.

Kim, J., Dubowitz, H., Hudson-Martin, E., & Lane, W. (2008). Comparison of 3 data collection methods for gathering sensitive and less sensitive information. *Ambulatory Pediatrics, 8,* 255–260.

Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., et al. (2001). Non-response and related factors in a nation-wide health survey. *European Journal of Epidemiology, 17,* 991–999.

Kroutil, L. A., Vorburger, M., Aldworth, J., & Colliver, J. D. (2010). Estimated drug use based on direct questioning and open-ended questions: Responses in the 2006 national survey on drug use and health. *International Journal of Methods in Psychiatric Research, 19,* 74–87.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47,* 2025–2047.

Kypri, K., Samaranayaka, A., Conner, J., Langley, J. D., & Maclennan, B. (2011). Non-response bias is a web-based health behavior survey of New Zealand tertiary students. *Preventive Medicine, 53,* 274–277.

Kypri, K., Stephenson, S., & Langley, J. (2004). Assessment of nonresponse bias in an internet survey of alcohol use. *Alcoholism, Clinical and Experimental Research, 28,* 630–634.

Lahaut, V. M. H. C. J., Jansen, H. A. M., van de Mheen, D., & Garretsen, H. F. L. (2002). Non-response bias in a sample survey on alcohol consumption. *Alcohol and Alcoholism, 37,* 256–260.

Lahaut, V. M. H. C. J., Jansen, H. A. M., van de Mheen, D., Garretsen, H. F. L., Verdurmen, J. E. E., & van Dijk, A. (2003). Estimating non-response bias in a survey of alcohol consumption: Comparison of response waves. *Alcohol and Alcoholism, 38,* 128–134.

Lamers, F., Hoogendoorn, A. W., Smit, J. H., van Dyck, R., Zitman, F. G., Nolen, W. A., et al. (2012). Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Comprehensive Psychiatry, 53,* 63–70.

Larsen, S. B., Dalton, S. O., Schüz, J., Christensen, J., Overvad, K., Tjønneland, A., et al. (2012). Mortality among participants and non-participants in a prospective cohort study. *European Journal of Epidemiology, 27,* 837–845.

Lavrakas, P. J. (2013). Applying a total error perspective for improving research quality in the social, behavioral, and marketing sciences. *Public Opinion Quarterly, 77,* 831–850.

Ledgerwood, D. M., Goldberger, B. A., Risk, N. K., Lewis, C. E., & Price, R. K. (2008). Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addictive Behaviors, 33,* 1131–1139.

Lemmens, P. H. H. M., Tan, E. S., & Knibbe, R. A. (1988). Bias due to non-response in a Dutch survey on alcohol consumption. *British Journal of Addiction, 83,* 1069–1077.

Leonard, K., Dunn, N. J., & Jacob, T. (1983). Drinking problems of alcoholics: Correspondence between self and spouse reports. *Addictive Behaviors, 8,* 369–373.

Levy, K. B., O'Grady, K. E., Wish, E. D., & Arria, A. M. (2005). An in-depth qualitative examination of the ecstasy experience: Results of a focus group with ecstasy-using college students. *Substance Use and Misuse, 40,* 1427–1441.

Lewin, T. (2002). Teenage drinking a problem but not in way study found. *New York Times*, February 27.

Lin, L. F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly, 59,* 236–258.

Link, M. W., & Mokdad, A. H. (2005). Effects of survey mode on self-reports of adult alcohol consumption: A comparison of mail, web and telephone approaches. *Journal of Studies on Alcohol, 6,* 239–245.

Livingston, M., Dietze, P. Ferris, J., Pennay, D., Hayes, L., & Lenton, S. (2013). Surveying alcohol and other drug use through telephone sampling: A comparison of landline and mobile phone samples. *BMC Medical Research Methodology*, 13.

Lowe, J. B., Windsor, R. A., Adams, B., Morris, J., & Reese, Y. (1986). Use of a bogus pipeline method to increase accuracy of self-reported alcohol consumption among women. *Journal of Studies on Alcohol, 47,* 173–175.

Luetgert, M. J., & Armstrong, A. H. (1973). Methodological issues in drug usage surveys: Anonymity, recency, and frequency. *The International Journal of the Addictions, 8,* 683–689.

Macera, C. A., Jackons, K. L., Davis, D. R., Kronenfeld, J. J., & Blair, S. N. (1990). Patterns of non-response to a mail survey. *Journal of Clinical Epidemiology, 43,* 1427–1430.

Mäkelä, P., & Huhtanen, P. (2010). The effect of survey sampling frame on coverage: The level of and changes in alcohol-related mortality in Finland as a test case. *Addiction, 105,* 1935–1941.

Malvin, J. H., & Moskowitz, J. M. (1983). Anonymous vs. identifiable self-reports of adolescent drug attitudes, intentions and use. *Public Opinion Quarterly, 47,* 557–566.

Martino, S. C., McCaffrey, D. F., Klein, D. J., & Ellickson, P. L. (2009). Recanting of life-time inhalant use: How big a problem and what to make of it. *Addiction, 104,* 1373–1381.

Mays, V. M., & Jackson, J. S. (1991). AIDS survey methodology with black Americans. *Social Science and Medicine, 33,* 47–54.

McAuliffe, W. E., Breer, P., Ahmadifar, N. W., & Spino, C. (1991). Assessment of drug abuser treatment needs in Rhode Island. *American Journal of Public Health, 81,* 365–371.

McCabe, S. E., Diez, A., Boyd, C. J., Nelson, T. F., & Weitzman, E. R. (2006). Comparing web and mail responses in a mixed mode survey in college alcohol use research. *Addictive Behaviors, 31,* 1619–1627.

McCabe, S. E., & West, B. T. (2016). Selective nonresponse bias in population-based survey estimates of drug use behaviors in the United States. *Social Psychiatry and Psychiatric Epidemiology, 51,* 141–153.

McCoy, T. P., Ip, E. H., Blocker, J. N., Champion, H., Rhodes, S. D., Wagoner, K. G., et al. (2009). Attrition bias in a U.S. internet survey of alcohol use among college freshmen. *Journal of Studies on Alcohol, 70,* 606–614.

McElrath, K. (2005). MDMA and sexual behavior: Ecstasy users' perceptions about sexuality and sexual risk. *Substance Use and Misuse, 40,* 1461–1477.

McKnight, C., Des Jarlais, D., Bramson, H., Tower, C., Abdu-Quader, A. J., Nemeth, C., et al. (2006). Respondent-driven sampling in a study of drug users in New York City: Notes from the field. *Journal of Urban Health, 83*(suppl. 1), 54–59.

Meiklejohn, J., Connor, J., & Kypri, K. (2012). The effect of low survey response rates on estimates of alcohol consumption in a general population survey. *PLOS One*, 7(4), e35527. Accessed August 7, 2016 at: http://dx.doi.org/10.1371/journal.pone.0035527)

Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly, 52,* 100–124.

Merachnik, D. (1972). Why initiate a drug survey? In S. Einstein & S. Allen (Eds.), *Proceedings of the First International Conference on Student Drug Surveys*. Baywood: Farmingdale, NY.

Merkle, D., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse*. New York, NY: Wiley.

Merline, A., Jager, J., & Schulenberg, J. E. (2008). Adolescent risk factors for adult alcohol use and abuse: Stability and change of predictive value across early and middle adulthood. *Addiction, 103,* 84–99.

Mewton, L., Slade, T., Teesson, M., Memedovic, S., & Krueger, R. F. (2014). Improving the diagnostic criteria for alcohol use disorders through survey methodology and cognitive interviewing. *International Journal of Methods in Psychiatric Research, 23,* 359–371.

Meyer, I. H., & Wilson, P. A. (2009). Sampling lesbian, gay, and bisexual populations. *Journal of Counseling Psychology, 56,* 23–31.

Midanik, L. T. (1982). The validity of self-reported alcohol consumption and alcohol problems: A literature review. *British Journal of Addiction, 77,* 357–358.

Midanik, L. T. (1994). Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: Results from the 1990 United States National Alcohol Survey. *Addiction, 89,* 407–412.

Midanik, L. T., & Greenfield, T. K. (2003). Telephone versus in-person interviews for alcohol use: Results of the 2000 National Alcohol Survey. *Drug and Alcohol Dependence, 72,* 209–214.

Midanik, L. T., Greenfield, T. K., & Rogers, J. D. (2001). Reports of alcohol-related harm: Telephone versus face-to-face interviews. *Journal of Studies on Alcohol, 62,* 74–78.

Midanik, L. T., & Hines, A. M. (1991). 'Unstandard' ways of answering standard questions: Protocol analysis in alcohol survey research. *Drug and Alcohol Dependence, 27,* 245–252.

Mieczkowski, T. (1989). The accuracy of self-reported drug use: An evaluation and analysis of new data. In N. A. Weiner & M. E. Wolfgang (Eds.), *Pathways to criminal violence*. Newbury Park, CA: Sage Publications.

Miller, P. V. (1997). Is 'up' right? The national household survey on drug abuse. *Public Opinion Quarterly, 64,* 627–641.

Miller, J. W., Gfroerer, J. C., Brewer, R. D., Maimi, T. S., Mokdad, A., & Giles, W. H. (2004). Prevalence of adult binge drinking: A comparison of two national surveys. *American Journal of Preventive Medicine, 27,* 197–202.

Moore, R. S., & Ames, G. M. (2002). Survey confidentiality vs. anonymity: Young men's self-reported substance use. *Journal of Alcohol & Drug Education, 47,* 32–41.

Morgan, M. J. (1999). Memory deficits associated with recreational use of "ecstasy" (MDMA). *Psychopharmacology (Berl), 141,* 30–36.

Morral, A. R., McCaffrey, D. F., & Chien, S. (2003). Measurement of adolescent drug use. *Journal of Psychoactive Drugs, 35,* 301–309.

Morral, A. R., McCaffrey, D., & Iguchi, M. Y. (2000). Hardcore drug users claim to be occasional users: Drug use frequency underreporting. *Drug and Alcohol Dependence, 57,* 193–202.

Muhib, F. B., Lin, L. S., Stueve, A., Miller, R. L., Ford, W. L., Johnson, W. D., et al. (2001). A venue-based method for sampling hard-to-reach populations. *Public Health Reports, 116,* 216–222.

Mulford, H. A., & Miller, D. E. (1959). Drinking in Iowa, I. Sociocultural distribution of drinkers. *Quarterly Journal of Studies on Alcohol, 20,* 704–726.

Murray, D. M., & Perry, C. L. (1987). The measurement of substance use among adolescents: When is the 'bogus pipeline' method needed? *Addictive Behaviors, 12,* 225–233.

Needle, R. H., McCubbin, H., Lorence, J., & Hochhauser, M. (1983). Reliability and validity of adolescent self-reported drug use in a family-based study: A methodological report. *International Journal of the Addictions, 18,* 901–912.

Nelson, D. E., Naimi, T. S., Brewer, R. D., & Roeber, J. (2010). US state alcohol sales compared to survey data, 1993–2006. *Addiction, 105,* 1589–1596.

Nelson, D. E., Powell-Griner, E., Town, M., & Kovar, M. G. (2003). A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *American Journal of Public Health, 93*, 1335–1341.

O'Malley, P. M., Bachman, J. G., & Johnston, L. D. (1983). Reliability and consistency in self-reports of drug use. *The International Journal of the Addictions, 18,* 805–824.

O'Malley, P. M., Johnston, L. D., Bachman, J. G., & Schulenberg, J. (2000). A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues, 30,* 35–54.

Osgood, D. W., Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1988). The generality of deviance in late adolescence and early adulthood. *American Sociological Review, 53,* 81–93.

Ouelett, L. J., Cagle, H. H., & Fisher, D. G. (1997). "Crack" versus "rock" cocaine: The importance of local nomenclature in drug research and education. *Contemporary Drug Problems, 24,* 219–237.

Owens, L., Johnson, T. P., & O'Rourke, D. (2001). Culture and item nonresponse in health surveys. In M. L. Cynamon & R. A. Kulka (Eds.), *Seventh Conference on Health Survey Research Methods*. DHHS Publication No. (PHS) 01-1013. Hyattsville, MD: National Center for Health Statistics.

Parrott, A. C., Lees, A., Garnham, N. J., Jones, M., & Wesnes, K. (1998). Cognitive performance in recreational users of MDMA or "ecstasy": Evidence for memory deficits. *Journal of Psychopharmacology, 12,* 79–83.

Paschall, M., & Freisthler, B. (2003). Does heavy drinking affect academic performance in college? Findings from a prospective study of high achievers. *Journal of Studies on Alcohol, 64,* 515–519.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.

Percy, A., McAlister, S., Higgins, K., McCrystal, P., & Thornton, M. (2005). Response consistency in young adolescent's drug use self-reports: A recanting rate analysis. *Addiction, 100,* 189–196.

Pernanen, K. (1974). Validity of survey data on alcohol use. In R. Gibbins, Y. Israel, H. Kalant, R. Papham, W. Schmidt, & R. Smart (Eds.), *Research advances in alcohol and drug problems* (Vol. 1). New York, NY: Wiley.

Perrine, M. W., Mundt, J. C., Searles, J. S., & Lester, L. S. (1995). Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *Journal of Studies on Alcohol, 56,* 487–490.

Petzel, T. P., Johnson, J. E., & McKillip, J. (1973). Response bias in drug surveys. *Journal of Consulting and Clinical Psychology, 40,* 437–439.

Plant, M. A., Chick, J., & Kreitman, N. (1980). The effects of response rates on levels of self-reported alcohol consumption and alcohol-related problems: Conclusions from a Scottish Study. *British Journal on Alcohol and Alcoholism, 15,* 158–163.

Plant, M. A., & Miller, T,-I. (1977). Disguised and undisguised questionnaires compared: Two alternative approaches to drinking behavior surveys. *Social Psychiatry*, *12*, 21–24.

Pleck, J. H., Sonenstein, F. L., & Ku, L. (1996). Black-white differences in adolescent males' substance use: Are they explained by underreporting by Blacks? *Journal of Gender, Culture, and Health, 1,* 247–265.

Poikolainen, K., & Kärkkäinen, P. (1983). Diary gives more accurate information about alcohol consumption than questionnaire. *Drug and Alcohol Dependence, 11,* 209–216.

Poikolainen, K., & Kärkkäinen, P. (1985). Nature of questionnaire options affects estimates of alcohol intake. *Journal of Studies on Alcohol, 46,* 219–222.

Poikolainen, K., Podkletnova, I., & Alho, H. (2002). Accuracy of quantity-frequency and graduated frequency questionnaires in measuring alcohol intake: Comparison with daily diary and commonly used laboratory markers. *Alcohol and Alcoholism, 37,* 573–576.

Popham, R. E., & Schmidt, W. (1981). Words and deeds: The validity of self-report data on alcohol consumption. *Journal of Studies on Alcohol, 42,* 355–358.

Poulin, C., MacNeil, P., & Mitic, W. (1993). The validity of a province-wide student drug survey: Lessons in design. *Canadian Journal of Public Health, 84,* 259–264.

Prause, J., Dooley, D., Ham-Rowbottom, K. A., & Emptage, N. (2007). Alcohol drinking onset: A reliability study. *Journal of Child & Adolescent Substance Abuse, 16,* 79–90.

Psaty, B. M., Cheadle, A., Koespsell, T. D., Diehr, P., Wickizer, T., Curry, S., et al. (1994). Race- and ethnicity-specific characteristics of participants lost to follow-up in a telephone cohort. *American Journal of Epidemiology, 140,* 161–171.

Ramirez-Valles, J., Kuhns, L. M., Campbell, R. T., & Diaz, R. M. (2010). Community involvement, stigmatized identities, and sexual risk in Latino sexual minorities. *Journal of Health and Social Behavior, 51,* 30–47.

Ramo, D. E., Liu, H., & Prochaska, J. J. (2012). Reliability and validity of young adults' anonymous online reports of marijuana use and thoughts about use. *Psychology of Addictive Behaviors, 26,* 801–811.

Ramstedt, M. (2010). How much alcohol do you buy? A comparison of self-reported alcohol purchases with actual sales. *Addiction, 105,* 649–654.

Reardon, M. L., Burns, A. B., Preist, R., Sachs-Ericsson, N., & Lang, A. R. (2003). Alcohol use and other psychiatric disorders in the formerly homeless and never homeless: Prevalence, age of onset, comorbidity, temporal sequencing, and service utilization. *Substance Use and Misuse, 38,* 601–644.

Rehm, J. (1998). Measuring quantity, frequency, and volume of drinking. *Alcohol: Clinical and Experimental Research, 22*, 4S–14S.

Rhem, J., Greenfield, T. K., Walsh, G., Xie, X., Robson, L., & Single, E. (1999). Assessment methods for alcohol consumption, prevalence of high risk drinking and harm: A sensitivity analysis. *International Journal of Epidemiology, 28,* 219–224.

Richter, L., & Johnson, P. B. (2001). Current methods of assessing substance use: A review of strengths, problems, and developments. *Journal of Drug Issues, 31,* 809–832.

Ridolfo, H. (2011). *Testing of the national HIV behavioral surveillance system: Results of interviews conducted 1/13/2011-4/5/2011.* Hyattsville, MD: Questionnaire Design Research Laboratory, National Center for Health Statistics, Centers for Disease Control and Prevention. Available from: http://wwwn.cdc.gov/qbank/report/Ridolfo_NCHS_2011_NHBSS%20HIV.pdf#page=43

Robbins, C., & Clayton, R. R. (1989). Gender-related differences in psychoactive drug use among older adults. *Journal of Drug Issues, 19,* 207–219.

Romelsjö, A. (1989). The relationship between alcohol consumption and social status in Stockholm. Has the

social pattern of alcohol consumption changed? *International Journal of Epidemiology, 18,* 842–851.

Room, R. (2007). Taking account of cultural and societal influences on substance use diagnoses and criteria. *Focus, 5,* 199–207.

Room, R., Janca, A., Bennett, L. A., Schmidt, L., & Sartorius, N. (1996). WHO cross-cultural applicability research on diagnosis and assessment of substance use disorders: An overview of methods and selected results. *Addiction, 91,* 199–220.

Rootman, I., & Smart, R. G. (1985). A comparison of alcohol, tobacco and drug use as determined from household and school surveys. *Drug and Alcohol Dependence, 16,* 89–94.

Safika, I., Johnson, T. P., & Levy, J. A. (2011). A venue analysis of predictors of alcohol use prior to sexual intercourse among female sex workers in Senggigi, Indonesia. *International Journal of Drug Policy, 22,* 49–55.

Satyanarayana, V. A., Vaddiparti, K., Chandra, P. S., O'Leary, C. C., Benegal, V., & Cottler, L. B. (2010). Problem drinking among married men in India: Comparison between husband's and wife's reports. *Drug and Alcohol Review, 29,* 557–562.

Schober, S., Caces, M.F., Pergamit, M., & Branden, L. (1992). Effects of mode of administration on reporting of drug use in the National Longitudinal Survey. In C. Turner, J. Lessler, & J. Gfoerer (Eds.), *Survey measurement of drug use: methodological studies*. DHHS Pub. No., ADM 92-1929. Rockville, MD: National Institute on Drug Abuse.

Schutz, C. G., & Chilcoat, H. D. (1994). Breach of privacy in surveys on adolescent drug use: A methodological inquiry. *International Journal of Methods in Psychiatric Research, 4,* 183–188.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93–105.

Serdula, M. K., Mokdad, A. H., Byers, T., & Siegel, P. Z. (1999). Assessing alcohol consumption: Beverage-specific versus grouped-beverage questions. *Journal of Studies on Alcohol, 60,* 99–102.

Sharma, A. K., Aggarwal, O. P., & Dubey, K. K. (2002). Sexual behavior of drug-users: Is it different? *Preventive Medicine, 34,* 512–515.

Shield, K. D., & Rehm, J. (2012). Difficulties with telephone-based surveys on alcohol consumption in high-income countries: The Canadian example. *International Journal of Methods in Psychiatric Research, 21,* 17–28.

Shillington, A. M., & Clapp, J. D. (2000). Self-report stability of adolescent substance use: Are there differences for gender, ethnicity and age? *Drug and Alcohol Dependence, 60,* 19–27.

Shillington, A. M., Clapp, J. D., & Reed, M. B. (2011a). The stability of self-reported marijuana use across eight years of the National Longitudinal Survey of Youth. *Journal of Child & Adolescent Substance Abuse, 20,* 407–420.

Shillington, A. M., Clapp, J. D., Reed, M. B., & Woodruff, S. I. (2011b). Adolescent alcohol use self-report stability: A decade of panel study data. *Journal of Child & Adolescent Substance Abuse, 20,* 63–81.

Shillington, A. M., Reed, M. B., Clapp, J. D., & Woodruff, S. I. (2011c). Testing the length of time theory of recall decay: Examining substance use report stability with 10 years of National Longitudinal Survey of Youth data. *Substance Use and Misuse, 46,* 1105–1112.

Shillington, A. M., Roesch, S. C., Reed, M. B., Clapp, J. D., & Woodruff, S. I. (2011d). Typologies of recanting of lifetime cigarette, alcohol and marijuana use during a six-year longitudinal panel study. *Drug and Alcohol Dependence, 118,* 134–140.

Shillington, A. M., Woodruff, S. I., Clapp, J. D., Reed, M. B., & Lemus, H. (2012). Self-reported age of onset and telescoping for cigarettes, alcohol, and marijuana: Across eight years of the National Longitudinal Survey of Youth. *Journal of Child & Adolescent Substance Abuse, 21,* 333–348.

Siddiqui, O., Mott, J. A., Anderson, T. L., & Flay, B. R. (1999). Characteristics of inconsistent respondents who have "ever used" drugs in a school-based sample. *Substance Use and Misuse, 34,* 269–295.

Simpura, J., & Poikolainen, K. (1983). Accuracy of retrospective measurement of individual alcohol consumption in men: A re-interview after 18 years. *Journal of Studies on Alcohol, 44,* 911–917.

Sims, M. (1969). *Comparison of sales figures of alcoholic beverages with types and amounts reported by canadian facts company, limited, in a market survey*, Toronto, CAN: Addiction Research Foundation, Substudy No. 1-29-69.

Single, E., Kandel, D., & Johnson, B. (1975). The reliability and validity of drug use responses in a large scale longitudinal survey. *Journal of Drug Issues, 5,* 426–443.

Slade, T., Teesson, M., Mewton, L., Mmedovic, S., & Krueger, R. F. (2013). Do young adults interpret the DSM diagnostic criteria for alcohol use disorders as intended? A cognitive interviewing study. *Alcoholism Clinical and Experimental Research, 37,* 1001–1007.

Smith, P. E., Remington, P. L., Williamson, D. F., & Anda, R. F. A. (1990). A comparison of alcohol sales data with survey data on self-reported alcohol use in 21 states. *American Journal of Public Health, 80,* 309–312.

Snow, D. L., Tebes, H. K., & Arthur, M. W. (1992). Panel attrition and external validity in adolescent substance use research. *Journal of Consulting and Clinical Psychology, 60,* 804–807.

Sobell, L. C., Agrawal, S., & Sobell, M. B. (1997). Factors affecting agreement between alcohol abusers' and their collaterals' reports. *Journal of Studies on Alcohol, 59,* 405–413.

Sobell, L. C., Cellucci, T., Nirenberg, T. D., & Sobell, M. B. (1982). Do quantity-frequency data underestimate

drinking-related health risks? *American Journal of Public Health, 72,* 823–828.

Sobell, L. C., & Sobell, M. B. (2003). Alcohol consumption measures. In J. P. Allen & V. B. Wilson (Eds.), *Assessing alcohol problems: A guide for clinicians and researchers, Second Edition.* NIH Pub. No. 03-3745. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Solowij, N., & Battisti, R. (2008). The chronic effects of cannabis on memory in humans: A review. *Current Drug Abuse Reviews, 1*, 81–98.

Steeh, C., Kirgis, N., Cannon, B., & DeWitt, J. (2001). Are they really as bad as they seem? Nonresponse rates at the end of the 20th Century. *Journal of Official Statistics, 17,* 227–247.

Stockwell, T., Donath, S., Cooper-Stanbury, M., Chiritzhs, T., Catalano, P., & Mateo, C. (2004). Under-reporting of alcohol consumption in household surveys: A comparison of quantity-frequency, graduated-frequency and recent recall. *Addiction, 99,* 1024–1033.

Stockwell, T., Zhao, J., Chiritzhs, T., & Greenfield, T. K. (2008). What did you drink yesterday? Public health relevance of a recent recall method used in the 2004 Australian National Drug Strategy Household Survey. *Addiction, 103,* 919–928.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stueve, A., & O'Donnell, L. N. (1997). Item nonresponse to questions about sex, substance use, and school: Results from the reach for health study of African American and Hispanic young adolescents. In J. H. J. Bancroft (Ed.), *Researching sexual behavior: Methodological issues*. Bloomington, IN: Indiana University Press.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The applications of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Swadi, H. (1990). Validating and improving the validity of self-reports in adolescent substance misuse surveys. *Journal of Drug Issues, 20,* 473–486.

Swaim, R. C., Beauvais, E. L., Chavez, E. L., & Oetting, E. R. (1997). The effect of school dropout rates on estimates of substance use among three racial/ethnic groups. *American Journal of Public Health, 87,* 51–55.

Tassiopoulos, K., Bernstein, J., Heeren, T., Levenson, S., Hingson, R., & Berstein, E. (2006). Predictors of disclosure of continued cocaine use. *Addictive Behaviors, 31,* 80–89.

Thrasher, J. F., Quah, A. C. K., Dominick, G., Borland, R., Driezen, P., Awang, R., et al. (2011). Using cognitive interviewing and behavioral coding to determine measurement equivalence across linguistic and cultural groups: An example from the International Tobacco Control Policy Evaluation Project. *Field Methods, 23,* 439–460.

Thygesen, L. C., Johansen, C., Keiding, N., Giovannucci, E., & Grønbæk, M. (2008). Effects of sample attrition in a longitudinal study of the association between alcohol intake and all-cause mortality. *Addiction, 103,* 1149–1159.

Tibblin, G. (1965). A population study of 50-year-old men: An analysis of the non-participation group. *Acta Medica Scandinavica, 178,* 453–459.

Tolonen, H., Laatikainen, T., Helakorpi, S., Talala, K., Martelin, T., & Prättälä, R. (2010). Marital status, educational level and household income explain part of the excess mortality of survey non-respondents. *European Journal of Epidemiology, 25,* 69–76.

Torvik, F. A., Rognmo, K., & Tambs, K. (2012). Alcohol use and mental distress as predictors of non-response in a general population health survey: The HUNT study. *Social Psychiatry and Psychiatric Epidemiology, 47,* 805–816.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tourangeau, R., Rips, L. J., & Raskinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60,* 275–304.

Tourangeau, T., Smith, T. W., & Rasinski, K. (1997). Motivation to report sensitive behaviors in surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology, 27,* 209–222.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 869–883.

Traugott, M., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43,* 359–377.

Trinkoff, A. M., & Storr, C. L. (1997). Collecting substance use data with an anonymous mailed survey. *Drug and Alcohol Dependence, 48,* 1–8.

Turner, C. F., Gribble, J. N., Al-Tayyib, A. A., & Chromy, J. R. (2002). Falsification in epidemiologic surveys: Detection and remediation (Prepublication Draft). *Technical Papers on Health and Behavior Measurement,* No. 53. Washington, D. C.: Research Triangle Institute. Available from: http://qcpages.qc.cuny.edu/~cturner/TechPDFs/53_Falsify.pdf

Turner, C. F., Lessler, J. T., & Devore, J. (1992). Effects of mode of administration and wording on reporting of drug use. In C. Turner, J. Lessler, & J. Gfroerer (Eds.), *Survey measurement of drug use: Methodological Studies*. DHHS Pub. No., ADM 92-1929. Rockville, MD: National Institute on Drug Abuse.

United States General Accounting Office. (1993). *Drug use measurement: Strengths, limitations, and recommendations for improvement, GAO/PEMD-93-18.*

Washington, D.C.: Program Evaluation and Methodology Division.

Valleroy, L. A., MacKellar, D. A., Karon, J. M., Rosen, D. H., McFarland, W., Shehan, D. A., et al. (2000). HIV prevalence and associated risks in young men who have sex with men. Young Men's survey study group. *Journal of the American Medical Association, 284,* 198–204.

Van Gorp, W. G., Wilkins, J. N., Hinkin, C. H., Moore, L. H., Hull, J., Horner, M. D., et al. (1999). Declarative and procedural memory functioning in abstinent cocaine abusers. *Archives of General Psychiatry, 56,* 85–89.

Van Loon, A. J. M., Tijhuis, M., Picavet, H. S. J., Surtees, P. G., & Ormel, J. (2003). Survey non-response in the Netherlands: Effects on prevalence estimates and associations. *Annals of Epidemiology, 13,* 05–110.

Vonmoos, M., Hulka, L. M., Prelier, K. H., Baumgartner, M. R., Stohler, R., Bolla, K. I., et al. (2013). Cognitive dysfunctions in recreational and dependent cocaine users: Role of attention-deficit hyperactivity disorder, craving and early age of onset. *British Journal of Psychiatry, 203,* 35–43.

Wagner, J., & Lee, S. (2015). Sampling rare populations. In T. P. Johnson (Ed.), *Handbook of health survey methods*. New York, NY: Wiley.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60,* 63–69.

Watten, R. G. (1996). Coping styles in abstainers from alcohol. *Psychopathology, 29,* 340–346.

Weinfurt, K. P., & Bush, P. J. (1996). Contradictory subject response in longitudinal research. *Journal of Studies on Alcohol, 57,* 273–282.

Weisner, C., Schmidt, L., & Tam, T. (1995). Assessing bias in community-based prevalence estimates: Towards an unduplicated count of problem drinkers and drug users. *Addiction, 90,* 391–405.

Weissman, A. N., Steer, R. A., & Lipton, D. S. (1986). Estimating illicit drug use through telephone interviews and he randomized response technique. *Drug and Alcohol Dependence, 18,* 225–233.

Welte, J. W., & Russell, M. (1993). Influence of socially desirable responding in a study of stress and substance abuse. *Alcoholism, Clinical and Experimental Research, 17,* 758–761.

Werch, C. E., Gorman, D. R., Marty, P. J., Forbess, J., & Brown, B. (1987). Effects of the bogus-pipeline on enhancing validity of self-reported adolescent drug use measures. *Journal of School Health, 57,* 232–236.

Whitehead, P., & Smart, R. (1972). Validity and reliability of self-reported drug use. *Canadian Journal of Criminology and Corrections, 14,* 1–8.

Wild, T. C., Cunningham, J., & Adlaf, E. (2001). Nonresponse in a follow-up to a representative telephone survey of adult drinkers. *Journal of Studies on Alcohol, 62,* 257–261.

Willis, G. (1997). The use of the psychological laboratory to study sensitive survey topics. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Wilson, P. (1981). Improving the methodology of drinking surveys. *The Statistician, 30,* 159–167.

Winters, K. C., Stinchfield, R. D., Henly, G. A., & Schwartz, R. H. (1991). Validity of adolescent self-report of alcohol and other drug involvement. *International Journal of the Addictions, 25,* 1379–1395.

Wish, E. D., Hoffman, J. A., & Nemes, S. (1997). The validity of self-reports of drug use at treatment admission and at follow-up: Comparisons with urinalysis and hair assays. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: improving the accuracy of survey estimates* (pp. 97–4147). Rockville, MD: NIDA Research Monograph 167, NIH Pub No, National Institute on Drug Abuse.

Witt, M., Pantula, J., Folsom, R., & Cox, C. (1992). Item nonresponse in 1988. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies.* Rockville, MD: National Institute on Drug Abuse.

Wolff, K., Farrell, M., Marsden, J., Monteiro, M. G., Ali, R., Welch, S., et al. (1999). A review of biological indicators of illicit drug use, practical considerations and clinical usefulness. *Addiction, 94,* 1279–1298.

Zanes, A., & Matsoukas, E. (1979). Different settings, different results? A comparison of school and home responses. *Public Opinion Quarterly, 43,* 550–557.

Zhao, J., Stockwell, T., & MacDonald, S. (2009). Non-response bias in alcohol and drug population surveys. *Drug and Alcohol Review, 28,* 648–657.

The Use of Biological Measures in Social Research on Drug Misuse

14

Michael Fendrich, Timothy P. Johnson, and Jessica Becker

14.1 Introduction

This chapter explores the role of biological testing for assessing and understanding the prevalence and nature of drug use and misuse in the community. It focuses less on the technological assessment of alternative methods than it does on the implementation and use of those methods. This is not meant to be an exhaustive review of the literature. Rather, it is intended to provide implementation examples, to evaluate the strengths and weaknesses of implementation, and to discuss and speculate on potential new areas of exploration for researchers examining substance use prevalence.

More specifically, this chapter focuses on the use of biological measures in "field" studies. Borrowing from Kerlinger and Lee (2000), field studies are defined as those that are done in real-life situations—and can include communities, schools, factories, organizations, and institutions. Excluded are methods and examples from clinical treatment settings, i.e., situations where those with substance dependence-related problems are being monitored because they are seeking treatment. Studies of "known" drug users were excluded (e.g., Colon et al. 2002). Also excluded are examples from settings where positive indicators of substance use may have direct and potentially adverse consequences for individuals being tested. Thus, although drug testing is widely used as a primary source of information about drug use in drug treatment and criminal justice settings, information gleaned from these contexts is not completely relevant to the purposes of this volume. Instead, biological measures for assessing illicit drugs, such as cocaine, marijuana, opiates, amphetamines, as well as for assessing legal drugs with a high potential for misuse (tranquilizers, prescription opiates), are explored in community settings. Although a great deal of recent research has addressed the development of biomarkers for heavy alcohol use (e.g., Berger et al. 2014), this review is limited to nonalcohol substances.

Over the past several decades, biological testing for drugs in community surveys has been used in three main ways. Many studies incorporate drug testing as an adjunct to self-report or interview measures in health and social surveys. The supplemental data provided by testing can be used to evaluate the quality of the responses provided by the survey informant. For example, by comparing informant reports and drug test results, researchers can tell the extent to which respondents accurately report their drug use. Aggregating positive test results in select

M. Fendrich (✉) · J. Becker
School of Social Work, University of Connecticut, West Hartford, CT, USA
e-mail: michael.fendrich@uconn.edu

J. Becker
e-mail: Jessica.becker@uconn.edu

T.P. Johnson
Survey Research Laboratory, College of Urban Planning and Public Affairs, University of Illinois at Chicago, Chicago, IL, USA
e-mail: timj@uic.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_14

samples, researchers have also used biological testing to estimate the prevalence of substance use in different social settings (e.g., dance clubs) and in entire communities. In the last decade, there have also been a spate of studies that have focused on estimating the comparative prevalence of drug use across different communities by testing for biomarkers in community sewage samples (e.g., Banta-Green et al. 2009).

14.2 Major Types of Biological Tests

A recent "White Paper" published by the American Society of Addiction Medicine (ASAM 2013) provides an extensive review of the history of drug testing as well as the properties of the major drug-testing "matrices" analyzed in drug tests. According to this report, drug testing began in the 1950s in hospital emergency room settings as a way to diagnose and treat patients admitted for overdose. Its use became widespread in drug treatment and criminal justice programs and in the military by the 1970s. In the 1980s, drug testing became widespread in the workplace, especially in settings involving the operation of motor vehicles, where safety was a primary concern. In the 1990s, drug testing was introduced into secondary schools. Most recently, there has been a push for increased use of testing to screen motor vehicle drivers suspected of driving while drunk or drugged. (ASAM 2013). The proliferation of testing coincides with rapid advancement in the development of testing technologies, an increased social acceptance of drug testing in multiple contexts, increased concerns about the negative impact of drugs on society, and supportive rulings by the U.S. Supreme Court, especially related to testing in the school context.[1]

Since drugs and their metabolites—the products of digestion of drugs—are distributed throughout the body, nearly any bodily fluid or "biological matrix" can be tested for drugs. Among the most common matrices used for drug testing are urine, oral fluid (saliva), hair, blood, and sweat. Both the "windows of drug detection" (the length of time after ingestion when a positive test result will be detected) and the degree of incorporation of drugs and metabolites vary by matrix type.

Figure 14.1, which was presented by Cone (2011) and published in the ASAM (2013) White Paper, depicts the variation in typical detection times across common matrices. Blood and oral fluid tend to have similar, short detection windows (typically one to two days). The detection times for urine are longer than blood and oral fluid and typically around three days. Hair testing has the longest detection time, typically 90 days (See Fig. 14.1). Nevertheless, as described in Verebey and Meenan (2011, p. 131), the detection times for urine, blood, and oral fluid vary by substance and by usage pattern. For example, short-acting barbiturates can be detected in the urine up to three days after last use. Marijuana can be detected in the urine for up to 30 days after last use for chronic marijuana users and only for three days for casual users. On the other hand, the window of detection for hair is 90 days, regardless of which substance is being tested (Verebey and Meenan 2011).

Of all drug-testing procedures, urine has been used most extensively in clinical and forensic contexts (Verebey and Meenan 2011). More recently, technical advances in oral fluid testing, along with its relative noninvasiveness, have increased its use (ASAM 2013; Verebey and Meenan 2011). Additionally, literature has

[1] *Numerous Supreme Court decisions have supported drug testing. For example, in* ***Vernonia School District v. Acton, 515 U.S. 646 (1995),*** *the Court upheld as* constitutional a school district policy which required students to consent to random drug testing as a condition for participation in interscholastic athletics. In ***Board of Education v. Earls,*** **122 S.Ct. 2559 (2002),** the Supreme Court held constitutional an Oklahoma school policy of randomly drug testing students who participate in competitive, nonathletic extracurricular activities, finding testing to be "a reasonably effective means of addressing the school district's legitimate concerns in preventing, deterring and detecting drug use." The latter ruling expanded the law, concluding drug testing to be reasonable for all extracurricular activities under the Fourth Amendment.

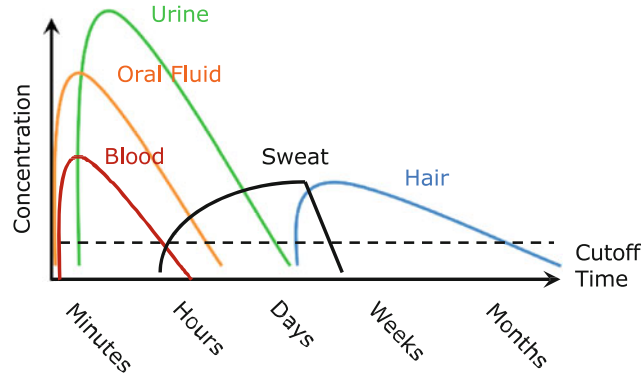

Fig. 14.1 Drug detection times in different matrices. From Cone (2011)

examined the properties of hair, fingernails, and sweat as repositories for drug metabolites (Verebey and Meenan 2011; Ropero-Miller et al. 2000). Challenges with respect to the possibility of environmental contamination affecting results and the difficult interpretation of dose–response relationships have undermined the utility of hair as a matrix, especially in forensic contexts. Nevertheless, these alternative matrices are especially attractive as "long term" indicators of substance use in epidemiological and survey research contexts where use detection has no clinical or legal ramifications.

In this chapter, studies employing biomarkers are classified into two broad categories: One category of studies, labeled here as "validity studies," contains those studies where biomarkers were used to confirm or validate data that was obtained from subject interviews, surveys, or self-reports. These studies typically employ benchmark comparison statistics, which are described in greater detail in the following summary. A second category of studies, branded as "prevalence studies," contains those where biomarker data often was used mainly to get an indicator of community use or prevalence, often in a stand-alone manner. In some prevalence papers, subject drug use reports were also obtained but the biological data was not used in any way to directly corroborate or evaluate individual reports of use.

This review excludes a growing and large body of US and international research focused on estimating maternal drug use among pregnant women (Friguls et al. 2012; Lester et al. 2001; Ostrea et al. 2001; Sanaullah et al. 2006). These studies typically collected maternal urine (Sanaullah et al. 2006) and hair (Friguls et al. 2012; Ostrea et al. 2001), and baby meconium (Lester et al. 2001; Ostrea et al. 2001) as well as maternal self-report. These studies were excluded due to noted inconsistencies with which comparison statistics were presented. For some studies, it was difficult to classify them as either validity or prevalence studies. More importantly, there are concerns that given the well-known teratogenic effects of drug exposure, maternal drug use reports in these studies are likely to be suppressed by perceptions regarding the social and legal consequences of disclosing drug use while pregnant. Although these studies are excluded here, they are worth a separate look in a stand-alone comprehensive review.

14.3 Validation Studies

In Table 14.1, 18 different studies were identified of nontreatment, noncriminal justice samples published between 1995 and 2014 that used biomarkers to validate self-reported drug use in some way. Samples varied widely in terms of age, but two-thirds (12 out of 18) focused on adults. Seven of the studies were focused on the "community," including special subgroups like MSM (men who have sex with men) and young adults, two focused on workers in the workplace, and four were either patients with a health condition (e.g., HIV positive) or at risk for disease. Specialized samples included in validation

studies included MSM (Fendrich et al. 2008), homeless women (Nyamathi et al. 2001), and African American adolescents and their parents (Delaney-Black et al. 2010). Two studies, both international, specifically focused on students: Basurto et al. (2009) examined students in Spain and Van Griensven et al. (2006) focused on students in Thailand.

In these 18 studies, three different types of biomarkers were employed—hair, urine, and oral fluid. Urine was the most commonly employed biomarker; it was used in 12 of the studies. Hair sampling was used in seven studies and oral fluid was used in only three studies. The specific drugs tested for varied considerably across studies. With three exceptions (Fendrich et al. 2004b; Fendrich et al. 2008; Hersch et al. 2002) most studies employed only one biomarker. Given biomarker variation in detection windows, this limits the extent to which comparisons with survey reports can be made.

With respect to drugs tested, cocaine was tested in 14 of the studies, marijuana in 11 of the studies, opioids (including heroin) in 13 of the studies, amphetamines in 8 studies, benzodiazepines in 3 studies and barbiturates, Ketamine, MDMA, and Rohypnol in only one study. Most studies tested for multiple drugs—although three studies, including two from outside the US, focused on only one substance. Abnet et al. (2004) examined adults at risk for esophageal cancer, focusing only on "opium," and van Griensven et al. (2006) examined students in Thailand, focusing only on amphetamines. Murphy et al. (2000) only studied urine for marijuana.

Validity studies showed a wide range of modes for assessing drug use in surveys. Most of the studies employed Audio Computer Self-Interviews (ACASI), a methodology widely accepted for the collection of drug use data in surveys (Office of Applied Studies 2001). Four studies were in-person interviews, four employed self-administered questionnaires (SAQ), two were computer-assisted telephone interviews (CATI), and two were phone interviews. There was one self-interview connected with a "palm pilot" (PASI) and one computer self-interview (CASI).

14.4 Statistics Used in Validity Studies

Since a goal of validation studies is to examine the quality of the self-report, certain benchmark comparative statistics are typically used in those studies. At a minimum, studies need to determine the consistency of the self-report with the biomarker results. Several main statistics are used for examining consistency. One of these is coefficient kappa (Cohen 1960), which is a measure of chance-corrected agreement. A related statistic is the "conditional kappa" (Bishop et al. 1975) which evaluates the same issue, conditional on a "positive" result (i.e., an affirmative self-report given that the biomarker indicates use of a drug). A third statistic is a simple indication of the overall percentage of agreement; that is, of all of the subjects tested, how many provided self-reports (positive or negative) about use that were consistent or concordant with the drug test result. The quality of the self-report may be further evaluated by examining self-report sensitivity (the percentage of those testing positive for a drug who disclose use of the drug) and specificity (the percentage of those testing negative for a drug who indicate that they have not used that drug). Finally, many reports include an overall prevalence statistic based on both the self-report and the biomarker test results.

Both the detection window for the biomarker and the reporting window for the self-report are of critical importance in understanding the validation process. Variation in biomarker detection windows can affect findings with regard to agreement. Take, for example, a respondent to a survey who used cocaine within 24 h of the time that the survey was administered. This respondent may not indicate as "positive" in the hair test for cocaine given the time necessary for drugs of abuse to metabolize and become incorporated into hair. This same respondent is likely to show a positive drug test result, regardless of his or her ongoing use patterns, if oral fluid test or urine tests were used. On the other hand, if the respondent is a frequent and chronic user of cocaine and has used the

Table 14.1 Validity studies incorporating biomarkers

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
Abnet et al. (2004)	Adults at risk for esophageal cancer-Iran	CATI	Urine	(150)	Opium	Lifetime	0.83	N/A	93	89	91	52.0	50.0
Basurto et al. (2009)	College students-Spain	SAQ	Urine	(506)	Cocaine	3 days	0.56	N/A	57	99	98	1.4	1.4
			Urine	(506)	Marijuana	3 days	0.66	N/A	92	90	89	22.1	7.9
Bharucha-Reid et al. (1995)	Workers	Phone	Urine	(162)	Amphetamines	12 months	0	N/A	N/A	N/A	N/A	1.9	0
			Urine	(162)	Barbiturates	12 months	−0.03	N/A	N/A	N/A	N/A	2.5	2.5
			Urine	(162)	Benzodiazepines	12 months	0.20	N/A	N/A	N/A	N/A	3.1	2.5
			Urine	(162)	Marijuana	12 months	0.63	N/A	70.6	95.2	N/A	11.7	10.5
			Urine	(162)	Cocaine	12 months	0.53	N/A	N/A	N/A	N/A	3.7	3.1
			Urine	(162)	Opiates	12 months	−0.02	N/A	N/A	N/A	N/A	3.1	1.2
Chen et al. (2006)	ER patients-Taiwan	In-person	Urine	(632)	Benzodiazepines	Currently	0.50	N/A	16	98	93	3.3	6.0
			Urine	(632)	Amphetamines	Currently	0.50	N/A	33	100	99	0.5	1.9
			Urine	(632)	Opiates	Currently	0.46	N/A	30	100	99	0.5	1.7
Colon et al. (2001)	Community-Puerto Rico	CATI	Hair	(114)	Cocaine	90 days	0.12	N/A	7	100	89	0.9	12.3
			Hair	(114)	Cocaine	Lifetime	0.18	N/A	14	98	88	3.5	12.3
			Hair	(114)	Heroin	90 days	0.49	N/A	33	100	98	0.9	2.6
			Hair	(114)	Heroin	Lifetime	0.66	N/A	67	99	98	2.6	2.6
Colon et al. (2010)	Community-Puerto Rico	ACASI	Urine	(523)	Marijuana	3 days	0.79	N/A	80	99	99	3.8	3.8
			Urine	(523)	Marijuana	12 months	0.43	N/A	95	92	92	11.9	3.8
			Urine	(523)	Marijuana	Lifetime	0.16	N/A	95	73	74	29.6	3.8

(continued)

Table 14.1 (continued)

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
			Urine	(523)	Cocaine	3 days	0.81	N/A	76	100	99	3.4	4.0
			Urine	(523)	Cocaine	12 months	0.61	N/A	91	94	96	7.5	4.0
			Urine	(523)	Cocaine	Lifetime	0.37	N/A	95	89	89	14.7	4.0
			Urine	(523)	Heroin	3 days	0.30	N/A	40	99	98	1.5	1.0
			Urine	(523)	Heroin	12 months	0.24	N/A	60	97	97	3.6	1.0
			Urine	(523)	Heroin	Lifetime	0.16	N/A	60	95	95	5.7	1.0
Delaney-Black et al. (2010)	Adolescents	N/A	Hair	(215)	Cocaine	Lifetime	0.02	N/A	N/A	N/A	N/A	0.9	33.7
	Parents/caregivers	N/A	Hair	(244)	Cocaine	Lifetime	0.29	N/A	N/A	N/A	N/A	6.1	28.3
			Hair	(248)	Opiates	Lifetime	0.16	N/A	N/A	N/A	N/A	3.3	7.0
Fendrich et al. (1999)	Community	ACASI	Hair	(322)	Cocaine	30 days	0.20	0.12	18	99	N/A	7.1	34.5
			Hair	(322)	Cocaine	12 months	0.17	0.10	19	95	N/A	9.6	34.5
			Hair	(322)	Cocaine	Lifetime	0.09	0.07	27	82	N/A	21.4	34.5
			Hair	(322)	Heroin	30 days	0.41	0.30	31	99	N/A	1.9	4.0
			Hair	(322)	Heroin	12 months	0.34	0.29	31	98	N/A	2.8	4.0
			Hair	(322)	Heroin	Lifetime	0.36	0.43	46	96	N/A	5.6	4.0
Fendrich et al. (2004b)	Community	ACASI	Oral fluid	(627)	Cocaine	12 months	0.23	0.36	40	95	N/A	6.8	5.7
			Oral fluid	(627)	Cocaine	3 months	0.24	0.27	30	97	N/A	4.2	5.7
			Oral fluid	(627)	Cocaine	30 days	0.34	0.28	30	99	N/A	2.7	5.7
			Oral fluid	(627)	Heroin	12 months	0.75	0.60	60	100	N/A	0.3	1.6
			Oral fluid	(627)	Heroin	3 months	0.75	0.60	60	100	N/A	0.3	1.6
			Oral fluid	(627)	Heroin	30 days	0.75	0.60	60	100	N/A	0.3	1.6
			Oral fluid	(623)	Marijuana	12 months	0.09	0.53	67	73	N/A	30.5	15.0

(continued)

Table 14.1 (continued)

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
			Oral fluid	(623)	Marijuana	3 months	0.11	0.51	62	79	N/A	23.3	15.0
			Oral fluid	(623)	Marijuana	30 days	0.11	0.41	52	83	N/A	20.3	15.0
			Hair	(627)	Cocaine	12 months	0.21	0.18	24	95	N/A	6.8	8.2
			Hair	(627)	Cocaine	3 months	0.25	0.18	22	98	N/A	4.2	8.2
			Hair	(627)	Cocaine	30 days	0.19	0.12	14	99	N/A	2.7	8.2
			Hair	(627)	Heroin	12 months	0.54	0.37	38	100	N/A	0.3	1.0
			Hair	(627)	Heroin	3 months	0.54	0.37	38	100	N/A	0.3	1.0
			Hair	(627)	Heroin	30 days	0.54	0.37	38	100	N/A	0.3	1.0
			Hair	(623)	Marijuana	12 months	0.02	0.65	75	72	N/A	30.5	0.6
			Hair	(623)	Marijuana	3 months	0.03	0.68	75	78	N/A	23.3	0.6
			Hair	(623)	Marijuana	30 days	0.02	0.39	50	82	N/A	20.3	0.6
			Urine	(627)	Cocaine	12 months	0.41	0.42	46	96	N/A	6.8	5.7
			Urine	(627)	Cocaine	3 months	0.39	0.32	35	98	N/A	4.2	5.7
			Urine	(627)	Cocaine	30 days	0.41	0.28	30	99	N/A	2.7	5.7
			Urine	(627)	Heroin	12 months	0.40	0.25	25	100	N/A	0.3	1.6
			Urine	(627)	Heroin	3 months	0.40	0.25	25	100	N/A	0.3	1.6
			Urine	(627)	Heroin	30 days	0.40	0.25	25	100	N/A	0.3	1.6
			Urine	(623)	Marijuana	12 months	0.46	0.73	80	81	N/A	30.5	15.0
			Urine	(623)	Marijuana	3 months	0.52	0.66	73	87	N/A	23.3	15.0
			Urine	(623)	Marijuana	30 days	0.58	0.65	71	91	N/A	20.3	15.0
Fendrich et al. (2008)	MSM	ACASI	Urine and oral fluid	(216)	Ketamine	30 days	0.32	0.20	23	N/A	N/A	3.2	10.6
			Urine and oral fluid	(216)	MDMA	30 days	0.16	1.00	100	N/A	N/A	5.6	0.7

(continued)

Table 14.1 (continued)

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
			Urine and oral fluid	(216)	Methamphetamine	30 days	0.66	0.59	60	N/A	N/A	3.7	3.6
			Urine and oral fluid	(216)	Rohypnol	30 days	N/A	N/A	N/A	N/A	N/A	0.0	0.8
			Urine and oral fluid	(216)	Cocaine	30 days	0.44	0.59	63	N/A	N/A	7.4	4.7
			Urine and oral fluid	(216)	Heroin	30 days	N/A	N/A	N/A	N/A	N/A	0.0	3.9
			Urine and oral fluid	(216)	Marijuana	30 days	.58	.65	74	N/A	N/A	25.5	22.7
Harrison et al. (2007)	Community-young adults (12–25 years)	ACASI	Urine	(3760)	Marijuana	30 days	0.54	N/A	67	93	90	12.6	11.1
			Urine	(3748)	Marijuana	7 days	0.59	N/A	58	97	93	9.0	11.1
			Urine	(3749)	Marijuana	3 days	0.60	N/A	54	98	93	7.5	11.1
			Urine	(3761)	Cocaine	30 days	0.24	N/A	23	99	98	1.1	1.4
			Urine	(3753)	Cocaine	7 days	0.26	N/A	19	100	99	0.7	1.4
			Urine	(3753)	Cocaine	3 days	0.26	N/A	16	100	99	0.5	1.4
			Urine	(3771)	Opiates	30 days	0.05	N/A	12	97	97	1.9	1.3
			Urine	(3754)	Opiates	7 days	0.02	N/A	3	99	98	1.1	1.3
			Urine	(3754)	Opiates	3 days	0.03	N/A	4	99	98	0.7	1.3
			Urine	(3762)	Amphetamines	30 days	0.08	N/A	10	98	98	0.6	1.1
			Urine	(3748)	Amphetamines	7 days	0.08	N/A	5	100	99	0.3	1.1
			Urine	(3748)	Amphetamines	3 days	0.08	N/A	5	100	99	0.2	1.1
Hersch et al. (2002)	Construction workers	SAQ	Hair	(229)	Marijuana	30 days	0.66	N/A	61	97	90	13.9	18.5
			Hair	(294)	Cocaine	30 days	0.55	N/A	43	99	94	4.9	9.9
			Hair	(229)	Marijuana	12 months	0.58	N/A	64	93	87	20.2	18.5

(continued)

Table 14.1 (continued)

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
			Hair	(294)	Cocaine	12 months	0.62	N/A	63	97	93	9.7	9.9
			Urine	(294)	Marijuana	30 days	0.62	N/A	88	93	93	13.9	9.9
			Urine	(294)	Cocaine	30 days	0.52	N/A	64	97	96	4.9	4.1
Johnson et al. (2009)	Young adults	SAQ	Oral fluid	(360)	Cocaine	2 days	0.45	N/A	33	100	91	4.4	12.8
			Oral fluid	(360)	Marijuana	2 days	0.53	N/A	68	92	86	12.7	9.7
			Oral fluid	(360)	Amphetamine	2 days	0.60	N/A	48	99	96	3.9	7.0
Ledgerwood et al. (2008)	Community	Phone/in-person	Hair	(538)	Marijuana	90 days	0.68	N/A	70	96	N/A	15.4	14.3
				(609)	Cocaine	90 days	0.61	N/A	100	94	N/A	4.8	10.3
				(613)	Opiates	90 days	0.55	N/A	44	100	N/A	2.6	1.5
				(611)	Methamphetamine	90 days	0.37	N/A	33	99	N/A	2.0	1.5
Murphy et al. (2000)	HIV-infected and high-risk youth	ACASI	Urine	(182)	Marijuana	2 days	N/A	0.57	N/A	N/A	64	17.0	15.0
Nyamathi et al. (2001)	Homeless women	In-person	Hair	(1037)	Cocaine	30 days	0.38	N/A	44	93	69	N/A	N/A
			Hair	(985)	Cocaine	4 months	0.53	N/A	64	88	77	N/A	N/A
			Hair	(1037)	Cocaine	6 months	0.50	N/A	68	82	75	42.0	49.0
			Hair	(1057)	Cocaine	12 months	0.48	N/A	73	75	74	N/A	N/A
			Hair	(1037)	Cocaine	Lifetime	0.38	N/A	83	76	69	N/A	N/A
			Hair	(114)	Heroin	Lifetime	0.66	N/A	67	99	98	2.6	2.6
Qian et al. (2014)	HIV-infected patients	CASI	Urine	(225)	Marijuana	7 days	0.68	N/A	70	94	86	28.6	36.0
			Urine	(225)	Cocaine	7 days	0.71	N/A	67	98	90	18.9	25.8
			Urine	(225)	Opiates	7 days	0.05	N/A	12	93	87	7.5	7.6
			Urine	(225)	Methamphetamine	7 days	−0.01	N/A	0	98	96	1.8	2.2

(continued)

Table 14.1 (continued)

Author	Sample	Survey mode	Biomarker	(*n*)	Drug	Reporting window	Kappa	*C*-Kappa	Sen (%)	Spec (%)	Agree (%)	Self-report prevalence (%)	Assay prevalence (%)
Van Griensven et al. (2006)	Students-Thailand	PASI	Urine	(328)	Amphetamines	7 days	0.38	N/A	40	98	96	3.1	3.1
		ACASI	Urine	(325)	Amphetamines	7 days	0.44	N/A	60	97	96	4.9	3.1
		SAQ	Urine	(310)	Amphetamines	7 days	0.32	N/A	29	99	97	1.6	2.3
		In-person	Urine	(317)	Amphetamines	7 days	0.66	N/A	75	99	99	1.6	1.3

N/A not available, Survey mode: *ACASI* audio-computer-assisted self-interview, *CASI* computer-assisted self-interview, *CATI* computer-assisted telephone interview, *PASI* palmtop-assisted self-interview, *SAQ* self-administered questionnaire, *C-Kappa* conditional kappa, *Sen* sensitivity of the self-report (biomarkers as the criterion), *Spec* specificity of the self-report (biomarker at the criterion), *Agree* agreement

substance multiple times in the past week (as well as in the past 24 h), he or she will show up as positive on the hair test even if there was no use in the past 24 h. As another example, if a subject last used cocaine within the past three months but not during the past month or past week, he or she is likely to generate a positive hair test result but a negative urine test result. Accordingly, depending exactly on when use occurred and on which type of biomarker was assessed, there may or may not be overlap between biomarkers and self-reports.

Epidemiological research on drug use typically assesses lifetime, past year, and past month use. Studies incorporating drug testing have added time frames that are more consistent with biomarker detection windows such as past 90-day use (corresponding to hair testing), and past 3-day use (corresponding to urine testing). In addition, there are studies reporting on past 7-day use (Harrison et al. 2007; Qian et al. 2014; van Griensven et al. 2006), past 2-day use (Johnson et al. 2009; Murphy et al. 2000), past 4-month use (Nymathi et al. 2001), and past 6-month use (Nymathi et al. 2001). It should be noted that one study (Chen et al. 2006) was not specific about the time frame in the assessment, using the term "currently" in their interview assessment to ascertain use of opiates, amphetamines, and benzodiazepines—this lack of specificity in assessment of the reporting period is problematic, not standard in the field and makes it difficult to interpret validity findings.

14.5 Findings

Use and values of comparison statistics were reviewed in the 18 validity studies. It is interesting to note that studies varied in their use of most of the agreement statistics, with three exceptions: Coefficient kappa or conditional kappa were presented (or able to be calculated from available data) in every study that we selected, as were self-report/survey and biomarker prevalence. Two researchers over four studies (Fendrich et al. 1999, 2004a, 2008;

Murphy et al. 2000) employed conditional kappa. Self-report sensitivity was calculated in all but three studies (Bharucha-Reid et al. 1995; Delaney-Black et al. 2010; Murphy et al. 2000) and self-report specificity was not presented in four studies (Fendrich et al. 2008; Bharucha-Reid et al. 1995; Delaney-Black et al. 2010; Murphy et al. 2000). Overall agreement statistics were omitted from six studies (Fendrich et al. 1999, 2004a, 2008; Bharucha-Reid et al. 1995; Delaney-Black et al. 2010; Ledgerwood et al. 2008).

Summary validity statistics by specific substance were also examined, focusing on the three most widely examined substances, cocaine, marijuana, and opiates. Also included was an "overall" summary that captures all substances examined in the 18 papers. The statistics are summarized in Table 14.2. The biomarker comparisons suggest a median kappa of 0.38, 0.58, and 0.40 for cocaine, marijuana, and opiates, respectively. The corresponding median values of sensitivity were 43, 71, and 38%. For specificity, the corresponding values were 98, 92, and 99%. As kappa is an estimate of reporting reliability, these findings suggest that marijuana reporting is the most reliable. These findings also support the notion that so-called "underreporting" is less salient for marijuana than it is for cocaine and opiates.

In total, there were 108 different comparisons contained in the validity studies described in Table 14.1. Of these, 104 comparisons facilitated calculation of prevalence differences between biomarkers and self-reports. For 42 of the comparisons (40%), the self-report prevalence was higher than the biomarker prevalence. For 56 of the comparisons (54%), the biomarker prevalence was higher than the self-report prevalence. For 6 comparisons (6%), the biomarker and self-report were exactly equal (66 within 3 points, 57 within 2 points). Although biomarkers tend to yield higher prevalence estimates than self-reports, these findings should be interpreted with caution as they could be affected by biomarker detection windows and survey question time frame. There is no overwhelming pattern suggesting that adding biomarkers has a consistent and marked impact on substance use prevalence estimates when they are used in conjunction with survey questions.

14.6 Prevalence Studies

Table 14.3 identifies and summarizes 34 different prevalence studies incorporating biomarkers to assess overall prevalence of drug use in various samples from the U.S. (11 or 31%) and over two dozen other countries (25 or 69%; note that several studies outside the U.S. included multiple countries; see below). Studies were published between 1993 and 2015. Oral fluid, which was collected in 16 prevalence studies, was the most common matrix for biomarker collection. An additional eight studies collecting biomarkers employed urine, seven studies employed blood, and four studies employed hair. This distribution of collection matrices underscores the notion that prevalence studies are typically focused on ascertaining current substance use.

This table included 12 of the 14 papers that were summarized in the recent systematic review of truck driver substance use studies by Girotto et al. (2014). These prevalence studies have different approaches, foci, and biomarkers as follows. Included are four club drug prevalence studies that focused on dance clubs where MDMA and related amphetamines are of primary interest (Arria et al. 2002; Gripenberg-Abdon et al. 2012; Miller et al. 2005, 2009). All of the club drug studies employed oral fluid testing. A total of 22 studies concentrated on sampling people in motor vehicles of any kind, including ten studies on people in passenger cars (Berning et al. 2015; Drummer et al. 2003, 2007; Gjerde et al. 2008, 2011; Kruger et al. 1995; Lacey et al. 2009; Reid et al. 2012; Senna et al. 2010; Wylie et al. 2005), and another 12 studies focused on truck drivers (Couper et al. 2002; Crouch et al. 1993; Gates et al. 2013; Gjerde et al. 2010, 2012; Labat et al. 2008; Leyton et al. 2012; Lund et al.

Table 14.2 Summary of validity statistics for major substances overall

	Number of studies	Median kappa	Range of kappa	Median sensitivity (%)	Range of sensitivity (%)	Median specificity (%)	Range of specificity (%)
Cocaine	14	0.38	0.02; 0.81	43	7; 100	98	75; 100
Marijuana	11	0.58	0.02; 0.79	71	50; 95	92	72; 99
Opiates	13	0.40	−0.02; 0.83	38	3; 93	99	89; 100
Overall[a]	18	0.4	−0.03; 0.83	50	0; 100	98	72; 100

[a]Includes all substances across 18 studies

1988; Mieczkowski 2010; Peixe et al. 2014; Silva et al. 2003; Yonamine et al. 2013). Among the nonmotor vehicle-related studies, three studies identified that they were focused on examining prevalence of drug use among workers in their jobs (Edvardsen et al. 2014; Gjerde et al. 2010; Tsanaclis et al. 2007). One study tested hair discarded at hair salons in Norway (Lund et al. 2013). The table includes five studies that examined sewage to assess drugs in the community (Banta-Green et al. 2009; Karolak et al. 2010; Reid et al. 2012; Thomas et al. 2012; Van Nuijs et al. 2011).

In most of the prevalence studies, the biomarker data are used as a stand-alone indicator of substance abuse. Accordingly, for most of the papers placed in this category, survey reports from individuals were not even collected. Club drug studies are an exception, however (Arria et al. 2002; Miller et al. 2005, 2009). For instance, in both Miller et al. (2005) and Miller et al. (2009), oral fluid biomarkers are combined with drug test results and combinations are sampled from subjects over time at the sampling venue. One reason this is done is that club drug users may not always have accurate information about the drugs they are consuming. Biomarkers facilitate a clearer picture of the specific type of pills that might have been ingested at the club or dance party.

Sewage studies are the only prevalence studies that do not sample individual people. Typically, sewage researchers derive data from influent samples at wastewater treatment plants (WWTPs) over a limited time period and, using toxicological analyses, document concentrations of illicit and pharmaceutical drugs in raw sewage (Banta-Green et al. 2009). Concentrations are compared across WWTPs located in different proximal regions that were sampled at similar time periods. Thus, for example, Banta-Green et al. (2009) compared different drug concentrations between WWTPs serving urban sites in Oregon with WWTPs serving rural sites. Similarly, Thomas et al. (2012) compared concentrations in sewage-treated WWTPs across 19 European cities. These studies report concentrations of various drugs and metabolites, not population prevalence. Although only five such studies are included in the table, there are a number of other similar studies in the literature (e.g., see Karolak et al. 2010). A major limitation of this approach to biomarker analysis is that since individuals are not directly sampled, there is no way of knowing how many people are actually using drugs in an area—only that there is some nonzero probability that drugs were consumed. Another issue is that certain drugs remain in effluent water even after treatment (Huerta-Fontela et al. 2008). It is not clear how the persistence of drugs in water may affect sewage estimates, even after population behavioral shifts in ingestion. Sewage concentrations of opioids may be elevated as a consequence of medically prescribed use and not of illicit use. In addition, elevated concentrations could be a consequence of greater use density in a particular area (a few people using a lot of drugs) and not of a greater number of drug users. Nevertheless, sewage studies are of growing interest, especially outside of the United States.

One of the challenges in providing summaries of prevalence findings (which is also relevant to a

Table 14.3 Prevalence studies incorporating biomarkers

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Arria et al. (2002)	Club drug—rave attendees in Baltimore–Washington area	Oral fluid	Ecstasy	18	48 h	20
Banta-Green et al. (2009)	Wastewater analysis—Oregon	Sewage	Cocaine, MDMA, methamphetamine	N/A	N/A	N/A
Berning et al. (2015)	Motor vehicle drivers—roadside surveys in the contiguous 48 states in the U.S.	Blood	THC	N/A	N/A	Weekend nighttime: 12.6
			Any illicit drug	N/A	N/A	Weekday daytime: 11.3 Weekend nighttime: 14.3
			Only medications	N/A	N/A	Weekday daytime: 10.3 Weekend nighttime: 6.9
		Oral fluid	Any illicit drug	N/A	N/A	Weekday daytime: 10.6 Weekend nighttime: 13.9
			Only medications	N/A	N/A	Weekday daytime: 8.4 Weekend nighttime: 5.9

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Couper et al. (2002)	Motor vehicle drivers—commercial tractor-trailer drivers in Oregon and Washington	Urine	Amphetamines	N/A	N/A	1.7
			Carboxy-tetrahydrocannabinols	N/A	N/A	4.3
			Cocaine and benzoylecgonine	N/A	N/A	1.1
			Diazepam	N/A	N/A	0.4
			Opiates and opioids	N/A	N/A	1.6
			Any illicit, prescription, or over-the counter drug	N/A	N/A	21
Crouch et al. (1993)	Motor vehicle drivers—fatally injured truck drivers in the U.S.	Blood	Amphetamine or methamphetamine	N/A	N/A	7
			Phenylpropanolamine, ephedrine, or pseudoephedrine	N/A	N/A	7
			Cannabinoids	N/A	N/A	13
			Cocaine and benzoylecgonine	N/A	N/A	8
			Any illicit drug			67
Drummer et al. (2003)	Motor vehicle drivers—drivers killed in Australian road traffic crashes	Blood	Benzodiazepines	N/A	N/A	4.1
			Cannabis	N/A	N/A	13.5
			MDMA	N/A	N/A	0.18
			Methamphetamine	N/A	N/A	1.5
			Opioids[a]	N/A	N/A	4.9
			Any illicit drug	N/A	N/A	26.7
Drummer et al. (2007)	Motor vehicle drivers—drivers in Victoria, Australia	Oral fluid	Methamphetamine	N/A	N/A	2.1
			MDMA	N/A	N/A	1.3
			THC	N/A	N/A	0.66

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Edvardsen et al. (2014)	Community—health professionals in Norway	Oral fluid	Amphetamine	N/A	N/A	0
			Benzodiazepines	N/A	N/A	3.1
			Cocaine	N/A	N/A	0.22
			THC	N/A	N/A	0.11
			MDMA	N/A	N/A	0
			Methamphetamine	N/A	N/A	0.11
			Opioids	N/A	N/A	1.0
			Any illicit drug	0	48 h	0.6
Gates et al. (2013)	Motor vehicle drivers—truck drivers in fatal crashes in the U.S.	Blood	Amphetamine	N/A	N/A	24.5
			Benzoylecgonine	N/A	N/A	15.3
			Cannabinoids	N/A	N/A	9.6
			Cocaine	N/A	N/A	17.3
			Methamphetamine	N/A	N/A	33.9
			Depressants	N/A	N/A	3.0
			Narcotics	N/A	N/A	5.0
			Any illicit drug	N/A	N/A	21.9
Gjerde et al. (2008)	Motor vehicle drivers—roadside survey in Norway	Oral fluid	Amphetamines	N/A	N/A	0.3
			Cocaine/benzoylecgonine	N/A	N/A	0.1
			MDMA	N/A	N/A	0
			Methamphetamine	N/A	N/A	0.12
			Morphine	N/A	N/A	0.08
			Tetrahydrocannabinol	N/A	N/A	0.6
			Any illicit drug	N/A	N/A	1.0

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Gjerde et al. (2010)	Community—truck drivers and employees in Norway	Oral fluid	Methamphetamine	N/A	N/A	0.4
			THC	N/A	N/A	1.0
			Medicinal drugs	4.2	48 h	2.7
			Any illicit drug	0.4	48 h	1.3
Gjerde et al. (2011)	Motor vehicle drivers—roadside survey in Norway	Oral fluid	Codeine	N/A	N/A	0.99
			Diazepam	N/A	N/A	0.73
			Tetrahydrocannabinol	0.5	N/A[b]	0.56
			Zopiclone	N/A	N/A	2.20
Gjerde et al. (2012)	Motor vehicle drivers—truck drivers in Norway	Oral fluid	Alprazolam	N/A	N/A	0.0
			Amphetamine	N/A	N/A	0.4
			Benzoylecgonine	N/A	N/A	0.4
			Clonazepam	N/A	N/A	0.1
			Cocaine	N/A	N/A	0.3
			Codeine	N/A	N/A	0.5
			Diazepam	N/A	N/A	0.1
			Lorazepam	N/A	N/A	0.0
			Methadone	N/A	N/A	0.1
			Methamphetamine	N/A	N/A	0.4
			MDMA	N/A	N/A	0.0
			Morphine	N/A	N/A	0.1
			Nitrazepam	N/A	N/A	0.1
			Oxazepam	N/A	N/A	0.2
			Tetrahydrocannabinol	N/A	N/A	1.5
			Zopiclone	N/A	N/A	2.6

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Gripenberg-Abdon et al. (2012)	Club drug—passengers on an electronic music dance event/cruise ship departing from Sweden	Oral fluid	Amphetamine	1.3	48 h	7.1
			Cannabis	2.1	48 h	1.5
			Cocaine	0.8	48 h	0.3
			Ecstasy/MDMA	0.5	48 h	2.5
			Methamphetamine	N/A	N/A	0.8
			Any illicit drug	3.7	48 h	10.1
Karolak et al. (2010)	Wastewater analysis—Paris, France area	Sewage	Amphetamine, buprenorphine, cocaine and benzoylecgonine, MDMA	N/A	N/A	N/A
Kapusta et al. (2006)	Community—men in Austria	Urine	Amphetamines	N/A	N/A	0.3
			Benzodiazepines	N/A	N/A	0.2
			Cocaine	N/A	N/A	0.4
			Opiates	N/A	N/A	2.7
			THC	N/A	N/A	5.1
			Any illicit drug	N/A	N/A	7.6
Kruger et al. (1995)	Motor vehicle drivers—roadside survey in Germany	Oral fluid	Amphetamines	N/A	N/A	0.08
			Barbiturates	N/A	N/A	0.6
			Benzodiazepines	N/A	N/A	2.7
			Cocaine	N/A	N/A	0.01
			Marihuana	N/A	N/A	0.6
			Opiates	N/A	N/A	0.7
			Any illicit drug	N/A	N/A	1

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Labat et al. (2008)	Motor vehicle drivers—truck drivers in France	Urine	Amphetamines	N/A	N/A	0.3
			Benzodiazepines	N/A	N/A	0.4
			Buprenorphine	N/A	N/A	1.8
			Cannabinoids	N/A	N/A	8.5
			Cocaine	N/A	N/A	0.1
			Methadone	N/A	N/A	0.5
			Opiates	N/A	N/A	4.1
			Any illicit drug	N/A	N/A	5
Lacey et al. (2009)	Motor vehicle drivers—roadside survey in the U.S.	Blood and/or oral fluid	Antidepressants	N/A	N/A	0.7
			Marijuana	N/A	N/A	6.9
			Narcotic analgesics	N/A	N/A	1.6
			Sedatives	N/A	N/A	0.9
			Stimulants	N/A	N/A	3.3
			Any illicit drug	N/A	N/A	11.3
Leyton et al. (2012)	Motor vehicle drivers—truck drivers in Sao Paulo, Brazil	Urine	Amphetamines	N/A	N/A	5.7
			Cannabinoids	N/A	N/A	1.1
			Cocaine	N/A	N/A	2.2
			Any illicit drug	N/A	N/A	9.3
Lund et al. (1988)	Motor vehicle drivers—tractor-trailer drivers in Tennessee	Blood and/or urine	Cannabinoids	N/A	N/A	15
			Cocaine metabolites	N/A	N/A	2
			Nonprescription stimulants[c]	N/A	N/A	12
			Prescription stimulants[d]	N/A	N/A	5
			Any illicit drug	N/A	N/A	29

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Lund et al. (2013)	Community—patrons of hair salons in Norway	Hair	Benzodiazepines and z-hypnotics	N/A	N/A	9.5
			Benzoylecgonine	N/A	N/A	1.0
			Cocaine	N/A	N/A	2.5
			Diazepam	N/A	N/A	2.5
			THC	N/A	N/A	3.0
			Methamphetamine	N/A	N/A	0.5
			MDMA	N/A	N/A	0.5
			Opioids	N/A	N/A	1.0
			Any illicit drug	N/A	N/A	5.0
Mieczkowski (2010)	Motor vehicle drivers—truck drivers and truck driver applicants in the U.S.	Hair	Amphetamine/methamphetamine	N/A	N/A	Applicants: 0.76 Employed: 0.14
			Cocaine	N/A	N/A	Applicants: 5.37 Employed: 2.12
			Heroin	N/A	N/A	Applicants: 0.06 Employed: N/A
			Marijuana	N/A	N/A	Applicants: 2.18 Employed: 0.35

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
			MDMA	N/A	N/A	Applicants: 0.03 Employed: N/A
			Opiates	N/A	N/A	Applicants: 0.26 Employed: N/A
			PCP	N/A	N/A	Applicants: 0.01 Employed: N/A
			Any illicit drug	N/A	N/A	Applicants: 9 Employed: 3
		Urine	Amphetamine/methamphetamine	N/A	N/A	Applicants: 0.12 Employed: N/A
			Cocaine	N/A	N/A	Applicants: 0.54 Employed: 0.27
			Heroin	N/A	N/A	Applicants: 0.01 Employed: N/A
			Marijuana	N/A	N/A	Applicants: 0.83

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
						Employed: 0.21
			Opiates	N/A	N/A	Applicants: 0.02 Employed: N/A
			PCP	N/A	N/A	Applicants: 0.02 Employed: N/A
			Any illicit drug	N/A	N/A	Applicants: 2 Employed: 0.6
Miller et al. (2005)	Club drug—patrons at electronic music dance events on East and West Coasts of the U.S.	Oral fluid	Cocaine	N/A	N/A	At entrance: 6.4[e] At exit: 8.1[e]
			Ecstasy	N/A	N/A	At entrance: 4.7[e] At exit: 7.5[e]
			Hallucinogens, sedatives, and opioids	N/A	N/A	At entrance: <5[e] At exit: <5[e]
			Marijuana	N/A	N/A	At entrance: 1/3[e] At exit: 15[e]

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
			Any illicit drug	N/A	N/A	At entrance: 45[e] At exit: 38[e]
Miller et al. (2009)	Club drug—patrons at electronic music dance events in San Francisco and Baltimore/D.C areas	Oral fluid	Amphetamines and stimulants[f]	N/A	N/A	At entrance: 6.9 At exit: 11.2
			Cocaine	N/A	N/A	At entrance: 12.2 At exit: 11.2
			Ketamine	N/A	N/A	At entrance: 0 At exit: 0.4
			Marijuana/hashish	N/A	N/A	At entrance: 12.7 At exit: 11.6
			Painkillers	N/A	N/A	At entrance: 0.8 At exit: 1.1
			Any illicit drug	N/A	N/A	At entrance: 24.4 At exit: 26.0
Peixe et al. (2014)	Motor vehicle drivers—truck drivers arriving at airport in Brazil	Urine	Amphetamines	8.1	30 days	1.6
			Cannabis	N/A	N/A	0

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
			Cocaine	N/A	N/A	4.8
			Any illicit drug	N/A	N/A	8.1
Reid et al. (2012)	Wastewater analysis—Norway	Sewage	Cocaine	N/A	N/A	N/A
	Motor vehicle drivers—drivers in Norway	Oral fluid	Cocaine	N/A	N/A	0.7
	Combined population and user-group survey	N/A	Cocaine	0.22	Estimated use "on a given day"	N/A
Senna et al. (2010)	Motor vehicle drivers—drivers suspected of being under the influence in Switzerland	Blood	Amphetamines	N/A	N/A	7
			Benzodiazepines	N/A	N/A	6
			Cannabinoids	N/A	N/A	48
			Cocaine and benzoylecgonine	N/A	N/A	25
			Methadone	N/A	N/A	5
			Opiates	N/A	N/A	10
			Medicinal drugs	N/A	N/A	8
			Any psychoactive substance (including alcohol)	N/A	N/A	89
Silva et al. (2003)	Motor vehicle drivers—truck drivers in Brazil	Urine	Amphetamine	N/A	N/A	4.8
			Methamphetamine	N/A	N/A	0
			Cannabinoids	N/A	N/A	0.28
			Cocaine	N/A	N/A	0.28
			Any illicit drug	N/A	N/A	5.63

(continued)

Table 14.3 (continued)

Author	Sample	Biomarker	Drug	Self-report prevalence (%)	Reporting window	Biomarker prevalence (%)
Thomas et al. (2012)[g]	Wastewater analysis—19 European cities in Belgium, Czech Republic, Spain, Finland, France, the UK, Croatia, Italy, the Netherlands, Norway, and Sweden	Sewage	Amphetamine, benzoylecgonine, cocaine, MDMA, methamphetamine, THC-COOH	N/A	N/A	N/A
Tsanaclis et al. (2007)	Community—individuals from different sectors (Medico-Legal, Workplace, Police, Clinical Monitoring, Schools, Research and Insurance) which use drug testing in the UK	Hair	Amphetamines	N/A	N/A	14[h]
			Cannabinoids	N/A	N/A	25[h]
			Cocaine	N/A	N/A	25[h]
			Opiates	N/A	N/A	26[h]
Van Nuijs et al. (2011)	Wastewater analysis—Brussels, Belgium	Sewage	Amphetamine, cocaine, MDMA, methadone, methamphetamine, heroin	N/A	N/A	N/A
Wylie et al. (2005)	Motor vehicle drivers—drivers in Scotland	Oral fluid	Cocaine	N/A	N/A	1.6
			MDMA	N/A	N/A	4.6
			THC	N/A	N/A	3.7
			Any illicit drug	N/A	N/A	16.8
Yonamine et al. (2013)	Motor vehicle drivers—truck drivers in Sao Paulo, Brazil	Oral fluid	Amphetamine/methamphetamine	N/A	N/A	0.64
			Cocaine	N/A	N/A	0.56
			THC	N/A	N/A	0.40
			Any illicit drug	1.2	N/A	1.68

N/A not available
[a]Includes morphine, codeine, methadone
[b]Estimated use per 24 hours based on self-reported use in the last 12 months
[c]Includes phenylpropanolamine
[d]Includes amphetamine
[e]Prevalence rates are based on both drug bioassays as well as self-report questionnaires
[f]Includes ecstasy
[g]Castiglioni et al. (2013) use the same data
[h]Prevalence rates for this study represent the median rate for 7 cities in the UK

lesser extent for the validity findings) is that studies focused on biomarkers typically vary in the extent to which they classify different substances that they detect. For example, studies testing for "stimulants" could include any variation of cocaine, MDMA, amphetamines, and methamphetamines in their definitions. The authors are not often consistent as to the metabolites they are using as criteria for a positive classification. Some authors report cocaine and benzoylecgonine grouped as positive indicators of cocaine use, while others provide separate test results for each metabolite. To make sense of these indicators for public health purposes, it is important for researchers to group the metabolites into common aggregate indicators that are consistently reported on behavioral surveys, that is, reporting on separate metabolites is not particularly useful. Consistent reporting of this information across studies would mark an important advance for the field.

Overall, the nature of the prevalence study samples showed extreme variation, which corresponded to the wide raging prevalence estimates. Thus, for example, the studies using oral fluid detected a range of cocaine use prevalence from 0.01 to 12.2% (with a median of 0.40%). Similarly, for marijuana, the oral fluid prevalence estimates ranged from 0.11 to 12.7% (with a median of 0.83%). Some of these samples involve those attending club drug venues, where prevalence would be expected to be elevated (e.g., Miller et al. 2005). Others involve samples where current drug involvement is unlikely, such as health professionals (Edvardson et al. 2014). Motor vehicle studies included persons who were identified and sampled from random traffic stops as well as those whose blood was tested after a fatal injury from a crash. Accordingly, it is difficult to interpret overall summaries of prevalence estimates for any substance for studies in the prevalence category. Nevertheless, the wide sample variation underscores the range of situations where biomarker sampling can be implemented to draw broad inferences about different communities and societal sectors.

14.7 Discussion

While behavioral reports by individuals may be critical for understanding drug use patterns and histories, this review suggests that biomarker assessment can be implemented in a variety of settings and populations—both with and without surveys. When used alongside surveys and with appropriate comparison statistics, biomarkers may provide valuable insights about the social context of drug use reporting accuracy. As suggested by the typical overlap between biomarker and survey findings, the use of biomarkers is not essential in a survey study unless the researcher has reason to believe that accuracy may be compromised in some way.

One set of concerns in employing biomarkers in drug use research relates to the overall feasibility of implementation. Feasibility was addressed in Fendrich et al. (2004a) where participation in each of three methods (hair, oral fluid, and urine) was compared in a community epidemiological study of adults. Hair testing typically involves cutting hair from the base of the scalp (hair from other parts of the body is not generally used due to its slower rate of growth). Accordingly, this procedure can be rejected for cosmetic reasons as well as for practical reasons (when respondents have insufficient hair or are bald). Urine testing may prove embarrassing in certain contexts, resulting in nonparticipation in this procedure. Indeed, when looking at overall rates of participation across the hair, oral fluid (saliva), and urine testing procedures, Fendrich et al. (2004a) found the highest rates of refusal for urine testing (23.7%), followed by hair (12%) and oral fluid (9.6%). For hair, 21.2% of respondents had insufficient hair to participate.

Other concerns are more practical. Biological testing requires a separate set of toxicological analyses, typically conducted off site by a laboratory. Assurances of laboratory quality are essential. For example, researchers in the US need to have assurances that laboratories used meet "CLIA" (i.e., Clinical Laboratory Improvement Amendments) standards and have

the appropriate CLIA certificate to operate. Researchers can also employ their own proficiency testing by randomly sending in known positive specimens to the labs engaged in the analysis. Employing outside laboratories to analyze specimens requires the researcher to establish systems to carefully track all specimen id numbers so that their analyzed results can be accurately merged with survey data (if there are planned validity comparisons). While this last point may seem obvious, multiple sources of data inevitably increase the potential for error.

In general, concerns that biomarker assessment will be resisted by subjects are unfounded, given the plethora of studies and samples where they have been implemented. When used on their own, biomarker assessment provides only limited information about community trends; this information needs to be informed by knowledge of biomarker chemistry—including detection windows and appropriate metabolites to look for. In order to make sense of the results, implementation of stand-alone biomarker studies also needs to be informed by a clear knowledge of the nature of the sample and the sampling process.

References

Abnet, C. C., Saddatian-Elahi, M., Pourshams, A., Boffetta, P., Feizzadeh, A., Brennan, P., et al. (2004). Reliability and validity of opiate use self-report in a population at high risk for esophageal cancer in Golestan, Iran. *Cancer Epidemiology, Biomarkers & Prevention, 13,* 1068–1070.

American Society of Addiction Medicine. (2013). *Drug testing: A white paper of the American society of addiction medicine (ASAM)*. North Bethesda, MD: ASAM. Retrieved from http://www.asam.org/magazine/read/article/2013/12/16/asam-releases-white-paper-on-drug-testing

Arria, A. M., Yacoubian, G. S., Fost, E., & Wish, E. D. (2002). Ecstasy use among club rave attendees. *Archives of Pediatrics and Adolescent Medicine, 156* (3), 295–296.

Banta-Green, C. J., Field, J. A., Chiaia, A. C., Sudakin, D. L., Power, L., & de Montigny, L. (2009). The spatial epidemiology of cocaine, methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) use: A demonstration using a population measure of community drug load derived from municipal wastewater. *Addiction, 104,* 1874–1880.

Basurto, Z. F., Montes, G. J. M., Cubos, F. P., Santed, S. F., Rios, L. F., & Moreno, M. A. (2009). Validity of the self-report on drug use by university students: Correspondence between self-reported use and use detected in urine. *Psicothema, 21,* 213–219.

Berger, L., Fendrich, M., Jones, J., Fuhrmann, D., Plate, C., & Lewis, D. (2014). Ethyl glucuronide in hair and fingernails as a long-term alcohol biomarker. *Addiction, 109,* 425–431.

Berning, A., Compton, R., & Wochinger, K. (2015). *Results of the 2013–2014 national roadside survey of alcohol and drug use by drivers*. Washington, DC: National Highway Traffic Safety and Administration, Office of Behavioral Safety Research. Retrieved from https://www.researchgate.net/publication/274718527_Results_of_the_2013-2014_National_Roadside_Survey_of_Alcohol_and_Drug_Use_by_Drivers

Bharucha-Reid, R., McCann, D., Schork, M. A., Foxman, B., Bass, A., Fraser, W., et al. (1995). A comparison of self-reported drug use with a urine drug screen in a working population. *Experimental and Clinical Psychopharmacology, 3,* 280–286.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: M.I.T. Press.

Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Hernádez, F., Reid, M., et al. (2013). Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. *Environmental Science and Technology, 47,* 1452–1460.

Chen, W. J., Fang, C.-C., Shyu, R.-S., & Lin, K.-C. (2006). Underreporting of illicit drug use by patients at emergency departments as revealed by two-tiered urinalysis. *Addictive Behaviors, 31,* 2304–2308.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Colón, H. M., Pérez, C. M., Meléndex, M., Marrero, E., Ortiz, A. P., & Suárez, E. (2010). The validity of drug use responses in a household survey in Puerto Rico: Comparison of survey responses with urinalysis. *Addictive Behaviors, 35,* 667–672.

Colón, H. M., Robles, R. R., & Sahai, H. (2001). The validity of drug use responses in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *International Journal of Epidemiology, 30,* 1042–1049.

Colón, H. M., Robles, R. R., & Sahai, H. (2002). The validity of drug use reports among hard core drug users in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *Drug and Alcohol Dependence, 67,* 269–279.

Cone, E. J. (2011). Oral fluid drug testing workshop: Pain management. *Society of forensic toxicology/the international association of forensic toxicologists (SOFT/TIAFT)*. September 25–30, 2011. San Francisco, CA.

Couper, F. J., Pemberton, M., Jarvis, A., Hughes, M., & Logan, B. K. (2002). Prevalence of drug use in

commercial tractor-trailer drivers. *Journal of Forensic Sciences, 47*(3), 562–567.

Crouch, D. J., Birky, M. M., Gust, S. W., Rollins, D. E., Walsh, J. M., Moulden, J. V., et al. (1993). The prevalence of drugs and alcohol in fatally injured truck drivers. *Journal of Forensic Sciences, 38,* 1342–1353.

Delaney-Black, V., Chiodo, L. M., Hannigan, J. H., Greenwald, M. K., Janisse, J., Patterson, G., et al. (2010). Just say "I don't": Lack of concordance between teen report and biological measures of drug use. *Pediatrics, 126,* 887–893.

Drummer, O. H., Gerostamoulos, D., Chu, M., Swann, P., Boorman, M., & Carins, I. (2007). Drugs in oral fluid in randomly selected drivers. *Forensic Science International, 190,* 105–110.

Drummer, O. H., Gerostamoulos, J., Batziris, H., Chu, M., Caplehorn, J. R. M., Roberston, M. D., et al. (2003). The incidence of drugs in drivers killed in Australian road traffic crashes. *Forensic Science International, 134,* 154–162.

Edvardesen, H. M. E., Karinen, R., Moan, I. S., Oiestad, E. L., Christophersen, A. S., & Gjerde, H. (2014). Use of alcohol and drugs among health professionals in Norway: A study using data from questionnaires and samples of oral fluid. *Journal of Occupational Medicine and Toxicology, 9*(8). Retrieved from http://occup-med.biomedcentral.com/articles/10.1186/1745-6673-9-8

Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. S., & Spiehler, V. (1999). Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests. *American Journal of Epidemiology, 149*(10), 955–962.

Fendrich, M., Johnson, T. P., Wislar, J. S., & Hubbell, A. (2004a). Drug test feasibility in a general population household survey. *Drug and Alcohol Dependence, 73,* 237–250.

Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A., & Spiehler, V. (2004b). The utility of drug testing in epidemiological research: Results from an ACASI general population survey. *Addiction, 99,* 197–208.

Fendrich, M., Mackesy-Amiti, M., & Johnson, T. P. (2008). Validity of self-reported substance use in men who have sex with men: Comparisons with a general population sample. *Annals of Epidemiology, 18,* 752–759.

Friguls, B., Joya, X., Garcia-Serra, J., Gómez-Culebras, M., Pichini, S., Martinez, S., et al. (2012). Assessment of exposure to drugs of abuse during pregnancy by hair analysis in a Mediterranean island. *Addiction, 107* (8), 1471–1479.

Gates, J., Dubois, S., Mullen, N., Weaver, B., & Bédard, M. (2013). The influence of stimulants on truck driver crash responsibility in fatal crashes. *Forensic Science International, 228,* 15–20.

Girotto, E., Mesas, A. E., Maffei, de Andrade, S. M., & Birolim, M. M. (2014). Psychoactive substance use by truck drivers: A systematic review. *Occupational and Environmental Medicine, 71*, 71–76.

Gjerde, H., Christophersen, A. S., Moan, I. S., Yttredal, B., Walsh, J. M., Normann, P. T., & Mørland, J. (2010). Use of alcohol and drugs by Norwegian employees: A pilot study using questionnaires and analysis of oral fluid. *Journal of Occupational Medicine and Toxicology, 5*(13). Retrieved from http://www.occup-med.com/content/5/1/13.

Gjerde, H., Christophersen, A. S., Normann, P. T., Pettersen, B. S., Sabaredzovic, A., Samuelsen, S. O., et al. (2012). Analysis of alcohol and drugs in oral fluid from truck drivers in Norway. *Traffic Injury Prevention, 13*(1), 43–48.

Gjerde, H., Normann, P. T., & Mørland, J. (2011). Can the use of psychoactive drugs in the general adult population be estimated based on data from a roadside survey of drugs and driving? *Norsk Epidemiology, 21,* 49–54.

Gjerde, H., Normann, P. T., Pettersen, B. S., Assum, T., Aldrin, M., Johansen, U., et al. (2008). Prevalence of alcohol and drugs among Norwegian motor vehicle drivers: A roadside survey. *Accident Analysis and Prevention, 40,* 1765–1772.

Gripenberg-Abdon, J., Elgán, T. H., Wallin, E., Shaafati, M., Beck, O., & Andréasson, S. (2012). Measuring substance use in the club setting: A feasibility study using biochemical biomarkers. *Substance Abuse Treatment, Prevention, and Policy, 7*(7). Retrieved from http://www.substanceabusepolicy.com/content/7/1/7.

Harrison, L. D., Martin, S. S., Enev, T., & Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population,* DHHS Pub. No. SMA 07-4249, Methodology Series M-7. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Hersch, R. K., McPherson, T. L., & Cook, R. F. (2002). Substance use in the construction industry: A comparison of assessment methods. *Substance Use and Misuse, 37,* 1331–1358.

Huerta-Fontela, M., Galceran, M. T., & Ventura, F. (2008). Stimulatory drugs of abuse in surface waters and their removal in a conventional drinking water treatment plant. *Environmental Science and Technology, 42,* 6809–6816.

Johnson, M. B., Voas, R. A., Miller, B. A., & Holder, H. D. (2009). Predicting drug use at electronic music dance events: Self-reports and biological measurement. *Evaluation Review, 33,* 211–225.

Kapusta, N. D., Ramskogler, K., Hertling, I., Schmid, R., Dvorak, A., Walter, H., et al. (2006). Epidemiology of substance use in a representative sample of 18-year-old males. *Alcohol and Alcoholism, 41,* 188–192.

Karolak, S., Nefau, T., Bailly, E., Solgadi, A., & Levi, Y. (2010). Estimation of illicit drugs consumption by wastewater analysis in Paris area (France). *Forensic Science International, 200,* 153–160.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Fort Worth, TX: Harcourt College Publishers.

Krüger, H-P., Schuylz, E., & Magerl, H. (1995). *The German roadside Survey 1992–1994. Saliva analyses from an unselected driver population: Licit and Illicit drugs*. Paper presented at the 13th International Conference on Alcohol, Drugs and Traffic Safety (T'95), Adelaide, Australia. Retrieved from http://druglibrary.org/schaffer/misc/driving/s3p3.htm

Labat, L., Fontaine, B., Delzenne, C., Doublet, A., Marek, M. C., Tellier, D., et al. (2008). Prevalence of psychoactive substances in truck drivers in the Nord-Pas-de-Calais region (France). *Forensic Science International, 174,* 90–94.

Lacey, J. H., Kelley-Baker, T., Furr-Holden, D., Voas, R. B., Romano, E., Ramirez, A., et al. (2009). *2007 national roadside survey of alcohol and drug use by drivers: Drug results*. Washington, DC: National Highway Traffic Safety and Administration, Office of Behavioral Safety Research.

Ledgerwood, D. M., Goldberger, B. A., Risk, N. K., Lewis, C. E., & Price, R. K. (2008). Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addictive Behaviors, 33,* 1131–1139.

Lester, B. M., ElSohly, M., Wright, L. L., Smeriglio, V. L., Verter, J., Bauer, C. R., et al. (2001). The maternal lifestyle study: Drug use by medconium toxicology and maternal self-report. *Pediatrics, 107,* 209–317.

Leyton, V., Sinagawa, D. M., Oliveira, K. C. B. G., Schmitz, W., Andreuccetti, G., De Martinis, B. S., et al. (2012). Amphetamine, cocaine and cannabinoids use among truck drivers on the roads in the state of Sao Paulo, Brazil. *Forensic Science International, 215,* 25–27.

Lund, A. K., Preusser, D. F., Blomberd, R. D., & Williams, A. F. (1988). Drug use by tractor-trailer drivers. *Journal of Forensic Sciences, 33*(3), 648–661.

Lund, H. M. E., Gjerde, H., de Courtade, S. M. B., Øiestad, E. K., & Christophersen, A. S. (2013). A Norwegian study of the suitability of hair samples in epidemiological research of alcohol, nicotine and drug use. *Journal of Analytical Toxicology, 37,* 362–368.

Mieczkowski, T. (2010). Urinalysis and hair analysis for illicit drugs of driver applicants and drivers in the trucking industry. *Journal of Forensic and Legal Medicine, 17,* 254–260.

Miller, B. A., Furr-Holden, D., Johnson, M. C., Holder, H., Voas, R., & Keagy, C. (2009). *Journal of Studies on Alcohol and Drugs, 70*(2), 261–268.

Miller, B. A., Furr-Holden, D., & Voas, K. B. (2005). Emerging adults' substance use and risky behaviors in club settings. *Journal of Drug Issues, 35*(2), 357–378.

Murphy, D. A., Durako, S., & Wilson, C. M. (2000). Marijuana use among HIV-positive and high-risk adolescents: A comparison of self-report through audio computer-assisted self-administered interviewing and urinalysis. *American Journal of Epidemiology, 152,* 805–813.

Nyamathi, A., Leake, B., Longshore, D., & Gelberg, L. (2001). Reliability of homeless women's reports: Concordance between hair assay and self-report of cocaine use. *Nursing Research, 50,* 165–171.

Office of Applied Studies. (2001). *Development of computer-assisted interviewing procedures for the national household survey on drug abuse*. Washington, DC: Substance Abuse and Mental Health Services Administration.

Ostrea, E. M., Knapp, D. K., Tannenbaum, L., Ostrea, A. R., Romero, A., Salari, V., et al. (2001). Estimates of illicit drug use during pregnancy by maternal interview, hair analysis, and meconium analysis. *Journal of Pediatrics, 138*(3), 344–348.

Peixe, T. S., de Almeida, R. M., Girotto, E., de Andrade, S. M., & Mesas, A. E. (2014). Use of illicit drugs by truck drivers arriving at Paranaguá Port Terminal, Brazil. *Traffic Injury Prevention, 15*(7), 673–677.

Qian, H-Z., Mitchell, V. J., Bebawy, S., Cassell, H., Perez, G., McGowan, C. C., et al. (2014). Current drug use and lack of HIV virologic suppression: Point-of-care urine drug screen versus self-report. *BMC Infectious Diseases, 14*(508). Retrieved from http://www.biomedcentral.com/1471-2334/14/508

Reid, M. J., Langord, K. H., Grung, M., Gjerde, H., Amundsen, E. J., Morland, J., & Thomas, K. V. (2012). Estimation of cocaine consumption in the community: A critical comparison of the results from three complimentary techniques. *BMJ Open* 2, e001637. Retrieved from http://bmjopen.bmj.com/

Ropero-Miller, J. D., Goldberger, B. A., Cone, E. J., & Joseph, R. E. (2000). The disposition of cocaine and opiate analytes in hair and fingernails of humans following cocaine and codeine administration. *Journal of Analytic Toxicology, 24,* 496–508.

Sanaullah, F., Gillian, M., & Lavin, T. (2006). Screening of substance misuse during early pregnancy in Blyth: An anonymous unlinked study. *Journal of obstetrics and Gynecology, 26*(3), 187–190.

Senna, M.-C., Augsburger, M., Aebi, B., Briellman, T. A., Donzé, N., Dubugnon, J.-L., et al. (2010). First nationwide study on driving under the influence of drugs in Switzerland. *Forensic Science International, 198,* 11–16.

Silva, O. A., Greve, J. M. D., Yonamine, M., & Leyton, V. (2003). Drug use by truck drivers in Brazil. *Drugs: Education, Prevention and Policy, 102*(2), 135–139.

Thomas, K. V., Bijilsma, L., Castiglioni, S., Covaci, A., Emke, E., Grabic, R., et al. (2012). Comparing illicit drug use in 19 European cities through sewage analysis. *Science of the Total Environment, 432,* 432–439.

Tsanaclis, L., & Wicks, J. F. C. (2007). Patterns in drug use in the United Kingdom as revealed through analysis of hair in a large population sample. *Forensic Science International, 170,* 121–128.

van Griensven, F., Naorat, S., Kilmarx, P. H., Jeeyapant, S., Manopaiboon, C., Chaikummao, S., et al. (2006). Palmtop-assisted self-interviewing for the collection of sensitive behavioral data: Randomized trial with drug use urine testing. *American Journal of Epidemiology, 163,* 271–278.

Van Nuijs, A. L. N., Mougel, J.-F., Tarcomnicu, I., Bervoets, L., Blust, R., Jorens, P. G., et al. (2011). Sewage epidemiology—A real-time approach to estimate the consumption of illicit drugs in Brussels, Belgium. *Environment International, 37,* 612–621.

Verebey, K. G., & Meenan, G. (2011). Diagnostic laboratory: Screening for drug abuse. In P. Ruiz, E. C. Strain, & J. H. Lowinson (Eds.), *Lowinson and Ruiz's substance abuse. A comprehensive textbook.* (5th ed., pp. 123–137). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Wylie, F. M., Torrance, H., Seymour, A., Buttress, S., & Oliver, J. S. (2005). Drugs in oral fluid Part II: Investigation of drugs in drivers. *Forensic Science International, 150,* 199–204.

Yonamine, M., Sanches, L. R., Paranhos, B. A. P. B., de Almeida, R. M., Andreuccetti, G., & Leton, V. (2013). Detecting alcohol and illicit drugs in oral fluid samples collected from truck drivers in the state of São Paulo, Brazil. *Traffic Injury Prevention, 14*(2), 137–131.

Part V
Special Populations

Conducting Research on Adolescent Substance Abuse

15

Dianne L. Kerr and Willie H. Oglesby

15.1 Introduction

Adolescents present a challenging population for substance abuse research. Epidemiological data confirm that they are at risk and a crucial population to study, yet difficulties arise in conducting research due to the illegality of substance use, problems accessing this population, and issues pertaining to obtaining consent for research purposes.

Adolescence is a time of experimentation with many health risk behaviors including alcohol, tobacco, and other drug (ATOD) use (McConnell et al. 2014). Stress and low self-esteem may be contributing factors to the initiation of substance abuse (Cornelius et al. 2014; Tavolacci et al. 2013; Marshal et al. 2013). Other characteristics of this age group may contribute as well. For example, adolescents often feel invulnerable regarding ATOD-related problems and consequences and believe "it will never happen to me." Many adolescents also have little thought for the future and what may happen as a consequence of what they are doing today. In addition, adolescence is normally a time of emotional volatility and there is typically a period of self-absorption where other's needs may not be perceived as important. Finally, adolescence is a time of maturation and growth of the body and the brain. It is also a time of risk taking which may be related to brain function. As a normal part of brain development, the limbic system develops on a steeper curve than the prefrontal cortex. This difference is greatest during adolescence, which may cause an imbalance in behaviors related to emotion and response to incentives rather than rational decision-making. This imbalance contributes to adolescent risk taking (Casey et al. 2008). For some, this is when substance use and addiction starts, and often continues into adulthood.

According to the National Institute on Alcohol Abuse and Alcoholism (NIAAA), teen drinking is linked to a higher lifetime risk of alcoholism, and teens that drink are likely to develop alcohol dependence at earlier ages and have chronic relapses of alcohol dependence throughout their lives (NIAAA 2006). In one study, researchers found nearly half of adults who were alcohol dependent started using as teens and met the criteria for alcoholism by the age of 21 (Hingson et al. 2006). Similarly, a national study by the National Center on Addiction and Substance Abuse (CASA) found that 9 out of 10 Americans who meet the medical criteria for addiction started ATOD use before the age of 18. In addition, 25% of Americans who began using addictive substances before age 18 were addicted as compared to 4% who began at age 21 or older. This higher addiction rate

D.L. Kerr (✉)
College of Education, Health, and Human Services, Kent State University, Kent, OH, USA
e-mail: dkerr@kent.edu

W.H. Oglesby
College of Population Health, Thomas Jefferson University, Philadelphia, PA, USA
e-mail: Billy.Oglesby@jefferson.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_15

among those who began using addictive substances before the age of 18 is attributed to the underdeveloped brain of youth this age. These adolescents are more likely to take risks that impair judgment, heighten the risk of addiction, and increase the use of addictive substances that interfere further with brain development. For this reason, CASA considers teen substance use "America's #1 public health problem" (CASA 2011).

The effect of substance abuse on brain maturation is an area of needed continued research. Research on youth brain development and marijuana use has recently been used by the White House to dispute a New York Times editorial recommending the legalization of marijuana in the U.S. (Green 2014). The negative effects of alcohol and marijuana use on youth brain development and function are supported by several research studies (Brown and Taper 2006; Squeglia et al. 2009; Witt 2010), although more research is needed on the etiology and consequences of adolescent drinking that integrates multiple levels (genetic, cellular, molecular, systems [neuroimaging], and behavioral) (Witt 2010). Further compounding the issue of youth ATOD use are the social aspects of adolescents' lives. Frequent changes in peer groups are common as adolescents seek to establish their identity and try to fit in. Adolescent peer groups are often supportive of alcohol and marijuana use in particular, and less often, other illicit drugs. This, combined with immature decision-making skills, may lead to substance abuse. Other social factors such as ATOD use by parents, siblings, and friends and tobacco and alcohol advertising targeted at youth may also contribute to substance abuse or addictive behaviors of adolescents (Office of Adolescent Health 2014).

In recent years, there has been a decrease in adolescent's use of certain illicit drugs such as methamphetamine, cocaine, hallucinogens, and LSD; however, marijuana use still remains an area of significant concern (USDHHS 2014) and has recently surpassed the use of tobacco among youth. Another area of concern is the nonmedical use of prescription medications among adolescents and young adults. Many adolescents believe that prescription drugs are safer to use than street drugs and often find them readily available in the family medicine cabinet or through the Internet. Adolescents commonly misuse pain relievers such as Vicodin and OxyContin (USDHHS 2014). Other prescription drugs commonly misused are the ADHD medications Adderall and Ritalin, also called "smart drugs". High school and college students, in particular, misuse these drugs, which are readily available from friends. The drugs are used to stay up all night to study or to increase concentration and focus. The downside is that these drugs are amphetamine-based, which makes them habit-forming, and the sale or use of them without a prescription is a felony (Trudeau 2009).

15.2 Special Populations of Adolescents and Substance Abuse Research

Many adolescents are at increased risk of substance abuse. These include runaway or "throwaway" youth, homeless youth and youth in juvenile detention or foster care. Many of these young people have been rejected by their families and are called "throwaway" youth who are told or asked to leave home or prevented from returning home by a parent or caregiver (National Resource Center on Domestic Violence 2013). This rejection may be due to substance abuse by the youth or their parents, youth identifying as lesbian, gay, bisexual, transgender, or questioning (LGBTQ), young women being pregnant, or a variety of other reasons. Many have run away to escape abusive situations such as family violence (including but not limited to physical, sexual, or emotional abuse or neglect) or parental substance abuse. Other causes of youth homelessness include poverty, lack of affordable housing and health care and systemic racism (National Resource Center on Domestic Violence 2013).

The National Alliance to End Homelessness (2015a, b) estimates that during a year over a half million unaccompanied, single youth up to age

24 experience a homelessness episode of longer than one week and approximately 380,000 of these youth are under the age of 18. The Alliance also estimates every year 110,000 of these homeless youth are LGBTQ and, as such, face even more risks than their heterosexual counterparts. Many homeless youth resort to "survival sex" and substance abuse. LGBTQ youth are particularly vulnerable to physical and sexual assault, often experience mental health problems, and are twice as likely to attempt suicide as their heterosexual peers. Mental health problems, self-harm, and suicide ideation and attempts are often attributed to minority stress due to the discrimination that these individuals face (Meyer 2003).

Research with homeless youth is challenging and limited. These youths present the major obstacle of lack of "parental consent" for research participation. Institutional Review Boards (IRBs) can waive parental consent when studies pose minimal risk and the requirement of obtaining permission is not reasonable such as in cases where the youth were homeless due to family abuse and attempts to receive parental consent may be a safety risk. Researcher attempts to gain parental consent may discourage youth from participation. Providing a detailed cover letter delineating potential risks and rationale for not obtaining consent may be helpful to the IRB in making a decision on conducting research with these vulnerable youths (Rew et al. 2000). Another issue is participant recruitment. Rew et al. (2000) found involving street outreach staff to invite youth to participate in their study resulted in 100% participation by homeless youth who were approached. Youth are often recruited for research at runaway and homeless shelters.

The link between substance abuse and criminal offending is well established (Chassin 2008). The National Institute on Drug Abuse (2006) found among adolescents detained for criminal offenses in the year 2000, over half of boys (56%) and 40% of girls tested positive for drug use. Thus, juvenile justice facilities are a major referral system for drug treatment for these youth. Sadly, this need for drug treatment and/or referral often goes unmet. Coordinated systems of care are needed to obtain the treatment that these young people need while they are incarcerated and after their release (Chassin 2008).

Conducting research with incarcerated youth presents its own unique set of challenges. Adolescents in juvenile detention are considered prisoners and research conducted with them must be in accord with the Guidance on the Involvement of Prisoners in Research of the Office for Human Research Protections of the Department of Health and Human Services. According to this guidance, approval for research on incarcerated youth requires special IRB compositions such as including someone who is a prisoner, has been a prisoner, or a prisoner representative who "has a close working knowledge, understanding, and appreciation of prison conditions from the perspective of the prisoner" (USDHHS 2003).

Another group of youth at greater risk of substance abuse is youth in foster care, particularly those "aging out" of the foster care system at age 18. These youths often do not have skills for self-sufficiency, financial literacy and resources, and career-related skills. In addition, many have cooccurring mental health and substance abuse problems. Some leave the foster care setting just prior to age 18 and return to a dysfunctional home setting with their parents, where family problems have not typically been resolved. Others run away from foster care settings and fall through the cracks (National Resource Center on Domestic Violence 2013). Research with youth aging out of the foster care system is sparse, although since they are over the age of 18 the parental consent issues no longer exist. More studies need to be conducted with these youths related to substance abuse and addiction, in order to determine how such issues may be better addressed and programs tailored to help them achieve a more promising future.

15.3 Survey Research and Preexisting Datasets

In order to determine the incidence and prevalence of substance use among youth, survey research in the form of questionnaires or

interviews is often conducted. Such survey research includes questionnaires such as the Youth Risk Behavior Survey, the National Survey of Drug Use and Health, and the Monitoring the Future Study. Additional national studies of youth ATOD use such as the CASA Columbia are also available through the National Center for Addiction and Substance Abuse (CASA 2011).

Many of these national survey research studies include data sets on their websites for use by researchers seeking to do secondary analyses. Secondary analyses are important, as investigating additional aspects of the data may lead to new findings. For example, the study of specific subgroups of the population (e.g., racial, ethnic, or sexual minority) as part of a preexisting data set may reveal new and valuable information about group differences or disparities. In addition, because substance use is a controversial topic in school settings, and researchers often have difficulties accessing school students, preexisting data may be a better solution than attempting to collect primary data in schools.

15.3.1 Youth Risk Behavior Survey

The Youth Risk Behavior Surveillance System (YRBSS) is funded by the Centers for Disease Control and Prevention (CDC) and employs the Youth Risk Behavior Survey (YRBS) to collect data in the school setting. The YRBS includes data on six categories of priority health risk behaviors including (1) tobacco use; (2) unhealthy eating; (3) inadequate physical activity; (4) alcohol and other drug use; (5) sexual behaviors that may result in HIV infection, other sexually transmitted diseases (STDs), and unintended pregnancy; and (6) behaviors that contribute to unintentional injury and violence. These behaviors, often established during youth, are identified as priority areas because they contribute to death, disability, and social problems in the United States (CDC 2011). Priority areas 1 and 4 address ATOD use specifically, although use of substances has also been linked to priority areas 5 and 6. The YRBS has been used for data collection since 1991. Typically administered to high school and at times middle school students, the most recent national YRBS results (2013) confirm that youth are at risk for the use of ATOD, even though their use has decreased over time for several substances. For example, the percentage of high school students using alcohol during the past 30 days decreased from 39% in 2011 to 34.9% during 2013, and binge drinking (5 or more drinks in a row) from 22% in 2011 to 20.8% in 2013, yet 1 in 3 students still report current alcohol use and 1 in 5 report current binge drinking. Marijuana use in the last 30 days increased slightly from 23% in 2011 to 23.4% in 2013 and is now more prevalent than all current cigarette use (15.7%) and all current tobacco use (cigarettes, smokeless tobacco, and cigars) combined (22.4%) (Kann et al. 2014). The most recent YRBS results (2015) were released in June 2016 for all high school students and in August 2016 for sexual minority high school students (CDC 2015a).

The YRBS is a public domain survey that can be used without permission. In addition, preexisting data from the YRBS can be downloaded for use by researchers from the following site: http://www.cdc.gov/healthyyouth/yrbs/data/. Frequently asked questions about using YRBS data are on the site to assist researchers. These can be found at the following web address: http://www.cdc.gov/healthyyouth/yrbs/faq.htm#analyzing.

15.3.2 National Survey of Drug Use and Health

The National Survey of Drug Use and Health (NSDUH) is an annual survey sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA). Randomly selected household interviews with individuals over the age of 12 help to determine ATOD behavior and respondents are given a $30 incentive for their participation. Although this survey is not limited to youth, it reports results by age groups including younger adolescents ages 12–17 and older adolescents and young adults ages 18–25. The NSDUH was conducted by the Research

Triangle Institute (RTI) from 1988 to 2014 (SAMHSA 2014). Beginning in August 2015 SAMHSA contracted with another vendor although ICPSR will continue to make data available for public use. The most recent (2014) NSDUH data is consistent with YRBS data in that alcohol, tobacco, and illicit drug use decreased among the 12–17 year olds in the past year. Marijuana use rates remained about the same among youth aged 12–17 (Center for Behavioral Health Statistics and Quality 2015).

SAMHSA shares data from the NSDUH with researchers through their site Substance Abuse and Mental Health Data Archive (SAMHDA). Researchers may download data in a variety of formats (e.g., SPSS, SAS, Stata, etc.) at the following link: http://www.icpsr.umich.edu/icpsrweb/SAMHDA/browse Data from the National Survey on Drug Use and Health, 2013 is currently available for access here: http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/35509 #datasetsSection.

The site also includes "quick tables" where researchers can select variables of interest and produce tables online using SAMHDA's online analysis tools. The NSDUH 2014 data is available for download at http://www.samhsa.gov/data/. The SAMHDA includes many data sets that can be downloaded for secondary analyses according to age groups. Recent data is also available for the Drug Abuse Warning Network (DAWN), which includes data on Emergency Department visits for drug-related problems: http://www.icpsr.umich.edu/icpsrweb/SAMHDA/quicktables.

15.3.3 National Addiction and HIV Data Archive Program

Another data resource for researchers is the National Addiction and HIV Data Archive Program (NAHDAP). NAHDAP offers preexisting data sets for researchers for secondary analyses on both substance abuse and HIV. NAHDAP recognizes that data collected for one purpose can often be used to pursue new and valuable lines of research (NAHDAP 2012). The site includes links to several youth data sets such as the CDC's National Youth Tobacco Survey of youth in grades 6–12; The Kids Count Data Center, the Health Resources and Services Administration's Data Warehouse, and for older adolescents/young adults, the American College Health Association's National College Health Assessment as well as other data sets. The site may be accessed via this link: http://www.icpsr.umich.edu/icpsrweb/NAHDAP/.

15.3.4 Monitoring the Future

The National Institute on Drug Abuse (NIDA) funds Monitoring the Future (MTF), an annual study of 8th, 10th and 12th graders conducted by researchers at the University of Michigan Institute for Social Research in Ann Arbor, Michigan. The MTF is supported by NIDA since 1975. In earlier years (1975–1991) MTF was administered just to high school seniors but in subsequent years, 8th and 10th graders were added. Monitoring the Future attempts to provide continued attention to substance use by youth and adults. Trends in substance use are analyzed over the years. For school-based surveys, parents are informed well in advance of the study so that they may choose to opt out their child (passive consent). The 2013 overview report provides details on all substances used by students in 8th, 10th, and 12th grades (Johnston et al. 2014).

Like the other national studies mentioned, the 2013 MTF study shows alcohol and tobacco use decreasing but the use of certain illicit drugs recently on the rise, mostly due to marijuana use. Importantly, perceived risk associated with marijuana use among all grades (8th, 10th, 12th) declined sharply. This lower perceived risk may lead to increases in marijuana use prevalence in the future (Johnston et al. 2014). The 2014 MTF results indicated the rates of cigarettes, alcohol, and prescription pain reliever use have declined. Even marijuana use appears to be leveling off, perhaps due to prevention efforts (NIDA 2014). Data from MTF study is available to researchers for secondary analysis on the NAHDAP website. Data may be downloaded from the following

site: http://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies?q=Monitoring+the+future+data&x=29&y=10.

The 2014 MTF survey for the first time included questions on the use of e-cigarettes. Researchers found that the use of e-cigarettes has now surpassed the use of traditional cigarettes among teens. Although most experts agree that e-cigarettes are less harmful than traditional cigarettes, they contain nicotine, which is potentially harmful to adolescent brain development (NIDA 2014). In addition, a recent report by the Centers for Disease Control and Prevention found an alarming increase in the use of e-cigarettes among middle school and high school teens in just one year (2013–2014). Middle school student use increased from 1.1% in 2013 to 3.9% in 2014 and high school student use increased from 4.5% in 2013 to 13.4% in 2014 (CDC 2015b). The American Lung Association expresses concern about the possible health effects of e-cigarettes and recommends that they be regulated by the Food and Drug Administration (American Lung Association 2016).

Other large national data sets are available that are not focused exclusively on substance use but have questions or question sets that may be analyzed individually pertaining to ATOD use. The National Longitudinal Study of Adolescent Health (Add Health) is an example. Investigators from the Add Health study interviewed a representative sample of 7th–12th grade students during the 1994–1995 school year. These youth were followed in a cohort for several years during which time four in-home interviews were conducted—the most recent of which was in 2008 when the youth were 24–32 years old. Datasets are available for secondary analysis for all four waves of the Add Health Study. A Wave V follow-up will take place in 2016–2018 as interviewers will track and trace participants to collect additional data and determine the emergence of chronic disease among this group. Although not specific to substance abuse, the Add Health study may be helpful for substance abuse researchers desiring to conduct secondary analysis on youth and substance abuse topics (UNC Carolina Population Center 2015). Add Health data is now available on the Maelstrom Research website at: https://www.maelstrom-research.org/mica/study/add-health.

15.4 Obstacles to Conducting Substance Abuse Research with Youth

Substance abuse research with adolescents presents many challenges because, by its very nature, substance abuse is controversial and illegal. Challenges include issues of assent, parental consent, recruitment, anonymity, and confidentiality. In order to assist researchers as well as institutional review boards and reviewers with such challenges, the National Advisory Council for Drug Abuse (NACDA) and the National Institute on Drug Abuse (NIDA) developed NACDA Guidelines for Substance Abuse Research Involving Children and Adolescents (NIDA 2012). These guidelines are not federal regulations but are advisory in nature.

15.4.1 Institutional Review Board Considerations

According to NACDA Guidelines, in order to achieve Institutional Review Board (IRB) approval, a proposal must meet one of the following four conditions: (1) the research must not involve greater than minimal risk, (2) the research may involve greater than minimal risk but provides the prospect of direct benefit to the individual, (3) the research involves greater than minimal risk and no prospect of direct benefit to the individual subjects, but is likely to yield generalizable knowledge about the subject's disorder or condition, or (4) the research is not otherwise approvable but presents an opportunity to understand, prevent, or alleviate a serious problem affecting the health or welfare of children. More information about these categories may be found in the NACDA Guidelines (Appendix 1; NIDA 2012). Those submitting proposals or applications for studies of youth

substance abuse for IRB approval should familiarize themselves with this document.

15.4.2 Assent, Consent, and Recruitment Issues

"Assent" refers to the youth's ability to agree to participate in research, while "consent" refers to the parent's permission for the youth to participate. Adolescent substance abuse research and consent issues are inextricably intertwined since students have typically not reached the age of majority (age 18 in most states) and often cannot be involved in research without the consent of at least one parent or guardian. According to NACDA, youth are typically able to assent at an intellectual age of seven (NIDA 2012). Thus, adolescents between 7 and 17 years of age can assent to participate in research. For example, the YRBS is conducted in middle schools and high schools with youth who are minors. As such, they can assent but will still need the permission of a parent or guardian to participate unless they are emancipated minors. Obtaining parental permission often presents a challenge to researchers interested in addressing this population.

In general, there are two types of permission in research studies: *passive consent* and *active consent*. In *passive consent,* parents/guardians are provided information about a research study and asked to sign and return a form provided by the researchers if they do *not* wish their child to be involved in the study. If the form is not returned an assumption is made that parents/guardians approve of their child's participation. This is also sometimes called "opt-out consent" where consent is assumed if the parent/guardian does not opt their child out of the research study by signing the form. Although this type of consent typically produces the best return rates, it is considered ethically questionable, as some parents may not even receive the information they need about the study or the consent form, particularly if it is sent home with their child. Even if parents receive the information, there is no guarantee that they have read or understand it (Macquarie University 2006). One of the most effective methods of informing parents about an upcoming research study in the schools is to send a letter home with the child's report card (Pokorny 2006). Researchers have suggested that Institutional Review Boards scrutinize all passive consent projects to determine if the procedure is ethical for the given sample (Range et al. 2001). This is especially important in projects involving children and adolescents in substance use research where this type of consent is usually not permitted.

In *active consent* situations, parents/guardians must sign a consent form approving their child's participation in the study. The child is not permitted to participate if the form is not returned to the school or researchers. This is also sometimes called "opt-in" consent where consent is not assumed unless the parent/guardian opts their child into the study by signing the form. Active consent typically yields lower numbers of youths participating in research due to parents' or guardians' failure to return the form while passive consent yields more participants. Passive consent is not typically suitable for controversial topics such as substance abuse. In fact, in studies that involve minors participating in illegal activity such as substance abuse, passive consent is often not permitted by schools or Institutional Review Boards (IRBs) (Pokorny 2006). In addition, many parents do not want their children questioned about substance use/abuse citing privacy concerns even though the surveys are anonymous. Alternatively, some parents want to have access to their children's survey responses, which also breeches the security of the study. Because of the controversies surrounding substance abuse research with adolescents, some parents and politicians have sought to shut down adolescent substance use research. Recently in Rapid City, Oklahoma the Rapid City Area Schools temporarily discontinued administering a youth behavioral survey, even though similar surveys have been administered there since 1989. The concern was that the survey administration would conflict with a new state law that stated that students are not required to take surveys questioning them on a variety of topics, one of which is "illegal activities" (Colias 2014).

Many times parental concerns are based on their own denial that their children are involved in such activities or fears that questioning youth about such issues will somehow give adolescents ideas to try illegal substances. There is no evidence that taking substance use surveys initiates the use of such substances among youth. Although sometimes students learn from taking surveys, changing behaviors typically requires a greater effort. There is also an issue of "minimal risk" that is particularly important to passive consent with substance use research. In general, if more than minimal risk is involved, active consent must be obtained, yet the definition of minimal risk is not clear. It is difficult to determine risk when one child may be upset by questions about ATOD use and another child may not be affected by such questions (Pokorny 2006).

Research has been conducted to determine best methods to maximize return rates for active parental consent (Wolfenden et al. 2009; Secor-Turner et al. 2010; O'Donnell et al. 1997). Wolfenden et al. (2009) conducted a review of several research databases for studies published between 1988 and 2008 to identify recruitment strategies for child research participants through schools. The authors identified several strategies for enhancing participation including promoting the research with principals, teachers, parents and students; direct contact with parents about study information via telephone or face-to-face; providing incentives to teachers and students; and having a member of the research team closely monitoring the recruitment process. Although these strategies are indicated for improving participation with a child, the same strategies may be effective with adolescents.

15.4.3 Anonymity and Confidentiality

Anonymity means that the subject's name is not known to the researcher or the subject's name cannot be linked to his or her data in any way. Sometimes in survey research the data are "de-identified" meaning any identifying characteristics are separated from the survey data (Lavrakas 2008). Because of the illegality of substance abuse behaviors among minors, anonymity of research responses must be assured, particularly if study research questions inquire about personal use behaviors. Such studies typically are given a higher level of review, such as a full board review by the IRB with the researchers present to answer questions and ensure the health and safety of the participants.

Confidentiality, on the other hand, is when the researchers know the subjects' names and can link their name to their responses, but keep the data private and out of the hands of people outside of the research team. Confidential data may only be released to third parties with the express consent of the individual from whom the data was gathered (Lavrakas 2008). Otherwise, because survey data can be attributed to an individual, special data security precaution must be taken and any analysis conducted must not allow for an individual to be identified. For example, in a qualitative research study with only a few subjects revealing personal use behaviors, subjects' names obviously cannot be associated with their answers. As is the case with most qualitative interview studies, pseudonyms should be created for subjects to maintain confidentiality. In such cases the investigator(s) should be the only one(s) with access to real names and these should be locked away. Confidentiality also requires that the subject should not be identifiable by any information that they may disclose. For example, if a study uses qualitative interviews of only a few individuals and interviews disclose unique circumstances of an individual's life, the subject may be able to be identified by those who know him/her by use of deductive reasoning. Efforts must be taken by researchers to avoid this type of situation.

15.5 Access Issues

Although most youth attend school, access to in-school youth for substance use research involves a variety of obstacles. These may include gaining access to the school and permission to conduct the study, concerns about the illegality of substance abuse in this population,

parental objections, limited time during the school day to administer questionnaires, and an emphasis on other areas of the curriculum pertinent to state proficiency exam preparation.

According to a study of 57 school administrators (principals and superintendents), administrators are interested in research that provides tangible benefits to the school, is consistent with their educational mission, is not burdensome, does not interfere with state assessment time or other busy times, and is credible and noncontroversial (Befort et al. 2008). Certainly substance abuse research with youth is controversial, which may deter school administrators from granting permission for their school to participate in such research. Providing incentives to schools may help with gaining access. This could be in the form of a dollar amount for each returned survey or even sharing findings and recommendations based on the survey results. State proficiency testing is one of the highest priorities for school administrators. Research that may benefit such testing (e.g., how substance use affects learning or grade point average) is valued. On the other hand, if research interferes with proficiency testing in any way, it is likely to be discontinued.

Many school administrators will not consent to research projects solely on substance use or abuse due to concerns regarding potential parental complaints. Those experienced in school-based research indicate even after such studies are approved, parental objections may bring a quick end to research projects already in progress. For this reason, the administration of adolescent ATOD questionnaires in community settings (e.g., YMCA, Boys or Girls Clubs, etc.) may be a preferred venue. Most youth involved in community programs are volunteers and not a "captive audience" as they would be in a school classroom. This would present fewer barriers to administering substance use surveys.

15.6 Summary/Conclusion

Substance abuse remains a serious problem among adolescents and young adults. Due to the controversial nature of substance abuse research with youth, it is important to follow all school district and community agency policies, clear the research with the appropriate IRBs, and utilize the ACDA guidelines. Youth assent and parental consent is needed to conduct research with minors and that consent may be active or passive, depending upon the specific circumstances of the study, the school's policies, and the IRB's recommendations. In order to avoid the lengthy delays and precarious nature of gaining access to youths in school, preexisting data sets are readily available for analysis from a number of government agencies. These data sets, many of which are nationally representative, can be a source of new knowledge when analyzed in new ways or with different subgroups.

Appendix 1

1.1 NACDA Guidelines for Substance Abuse Research Involving Children and Adolescents[1]

I. Preamble

The National Advisory Council on Drug Abuse (NACDA) recognizes that substance abuse research involving children and adolescents is vital to understanding factors contributing to the initiation, maintenance and cessation of substance use and abuse among this population. This period of life is characterized by growth and maturation of brain and body, which potentially affects responses to drugs and treatment. Moreover, the great majority of people who develop substance use disorders (SUDs) or addiction begin to use drugs when they are young. Therefore, study of this population is crucial in order to develop effective prevention and treatment interventions, both behavioral and pharmacological, for youth.

Research on substance abuse involving children/adolescents should be designed,

[1]The document is available via the web at: https://www.drugabuse.gov/sites/default/files/pdf/nacdaguidelines.pdf. Accessed June 2016.

reviewed and conducted within the broader ethical principles outlined in the Belmont Report (discussed in greater detail in the NACDA Guidelines for Administration of Drugs to Human Subjects; (https://www.drugabuse.gov/funding/clinical-research/nacda-guidelines-administration-drugs-to-human-subjects) and the Code of Federal Regulations 45 CFR Part 46 Subpart A, and the additional protections for children under Subpart D. The reader is also referred to the general guidelines that have been developed specifically for the pediatric population: Guidelines for Ethical Conduct of Studies to Evaluate Drugs in Pediatric Populations (RE9503), American Academy of Pediatrics, Committee on Drugs (http://pediatrics.aappublications.org/content/95/2/286); Ethical Standards for Research with Children, Society for Research on Child Development (http://www.srcd.org/about-us/ethical-standards-research); Institute of Medicine, The Ethical Conduct of Clinical Research Involving Children; and Shah et al. (2004) How do Institutional Review Boards Apply the Federal Risk and Benefit Standards for Pediatric Research, JAMA, 29 (4), 476–481.

II. Purpose of These Guidelines

Research on substance use and abuse among children and adolescents presents its own unique challenges. As a result, the National Advisory Council for Drug Abuse (NACDA) and the National Institute on Drug Abuse (NIDA) have developed these guidelines to assist researchers, institutional review boards and study reviewers in developing, conducting or reviewing studies involving children and adolescents. The guidelines that are provided in this document address both general issues regarding conducting research in youth as well as issues that may specifically arise when conducting drug abuse research in youth. These guidelines are not codified and do not constitute Federal regulation. These guidelines are not intended to supplant the functions of either the Institutional Review Board (IRB) or the Office for Human Research Protections (OHRP). They are advisory to applicants, IRBs, Integrated Review Groups (IRGs), and others.

III. General Issues

The NACDA recommends consideration of a number of general issues applicable to studying substance use and abuse in children/adolescents. These issues are:

A. Federal Regulations for conducting research involving children

The regulations require that an Institutional Review Board (IRB) reviewing research involving children as subjects consider "the risks of harm or discomfort inherent in the proposed research and the anticipated benefits to the child subjects or society in general (OHRP 2001)." The regulations do not include as research risks any risks the child would be exposed to as part of clinical care. Children are defined as "persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research will be conducted" (45 CFR 46.402). For most states, under most conditions, this legal age is 18 years old[2] (for exceptions, see section entitled "Consent from Minors").

To receive IRB approval, the proposed research must fall into one of four categories.

1. Research not involving greater than **minimal risk** (45 CFR 46.404).
2. Research involving greater than minimal risk but presenting the **prospect of direct benefit** to the individual subjects (45 CFR 46.405) if the IRB finds that:

   (a) The risk is justified by the anticipated benefit to the subjects;
   (b) The relation of the anticipated benefit to the risk is at least as favorable to the

[2]Please note that for NIH-funded clinical research, children are defined as individuals "under the age of 21"; however individuals between the ages of 18 and 21 are permitted to consent to participate in research.

subjects as that presented by alternative approaches; and
(c) Adequate provisions are made for soliciting the assent of the children and permission of their parents or guardians.

3. Research involving greater than minimal risk and no prospect of direct benefit to the individual subjects, but likely to yield generalizable knowledge about the subject's disorder or condition (45 CFR 46.406). Research in this category is approvable provided:

   (a) The risk represents a **minor increase over minimal risk**;
   (b) The research intervention/procedure presents experiences to subjects that are reasonably commensurate with those inherent in their actual or expected medical, dental, psychological, social or educational situations; and
   (c) The intervention/procedure is likely to yield generalizable knowledge about the subjects' **disorder or condition**, which is of vital importance for the understanding or amelioration of the subjects' disorder, or condition.

4. Research that is not otherwise approvable which presents an opportunity to understand, prevent, or alleviate a serious problem affecting the health or welfare of children (45 CFR 46.407). Research that the IRB finds does not meet the requirements of 45 CFR 46.404, 46.405, 46.406, may be supported by DHHS provided:

   (a) The IRB finds the research presents a reasonable opportunity to further understanding, prevention, or alleviation of a serious problem affecting the health or welfare of children; and
   (b) The Secretary, after consultation with a panel of experts and following an opportunity for public review and comment, determines that the research satisfies one of the 45 CFR 46.404, 46.405, or 46.406 categories or the research presents a reasonable opportunity to further understanding, prevention, or alleviation of a serious problem affecting the health or welfare of children, and will be conducted in accordance with sound ethical principles, and adequate provisions are made for soliciting the assent of children and the permission of their parents or guardian.

Minimal risk (see Box)
Minimal risk is defined as the level of risk where "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests (45 CFR 46.102.i)." General consensus in the literature is that the risks of daily life refer to the daily lives of **normal, average, healthy** children living in safe environments and should be considered using rational means.

Minor increase over minimal risk
A minor increase over minimal risk is not defined in the regulations. "The Office for Human Research Protections, Department of Health and Human Services, believes that it is an appropriate responsibility of the IRB to determine when research would involve a minor increase over minimal risk" (Stith-Coleman, OHRP, personal communication). The statement in the regulations that "the research intervention/procedure presents experiences to subjects that are reasonably commensurate with those inherent in their actual or expected medical situations" suggests that "minor increase over minimal risk" could be a relative standard, i.e., a minor increase over the risk that a child with this "condition" is exposed to in daily life. While no guidelines have been issued by OHRP on whether or not "minor increase over minimal risk" should be an absolute or relative standard, both the IOM (2004) and the Secretary's Advisory Committee on Human Research Protections (SACHRP) have recommended that a minor increase over minimal

risk be considered an absolute standard similar to that for minimal risk.

Prospect of direct benefit

Prospect of direct benefit is not defined in the regulations. Direct benefit is usually considered to be medical/psychological benefits from research procedures only. Payments for participation in research or added psychological or medical interventions should not be considered a benefit.

Disorder or condition

The definition of a disorder or condition is not specified in the regulations. The recent IOM report states that limiting the definition of "disorder or condition" to an illness, disease or injury would result in too narrow a definition, whereas broad interpretations of any social, developmental or other characteristic, could unjustly single out groups of children already burdened by social disadvantages for research that would not necessarily benefit them. The IOM therefore recommended that "the term 'condition' should refer to a specific physical, psychological, neurodevelopmental or social characteristic that an established body of scientific evidence or clinical knowledge has shown to negatively affect children's health and well-being or to increase their risk of developing a health problem in the future." Therefore, given available scientific evidence, a "condition" may include risk factor (s) associated with a disorder that differentiate individuals from the general population. For example, because studies have demonstrated that exposure to trauma increases an individual's risk for substance abuse; children exposed to trauma may be considered to have a "condition" predisposing them to substance abuse.

B. Participant informed consent/assent

Assent is defined as a "child's affirmative agreement to participate in research" and applies to children whom the IRB judges to be capable of providing assent. General consensus appears to be those children who have reached an intellectual age of 7 years old, however a precise age has not been specified in the regulations. Therefore, the IRB must make a determination about the appropriate age for obtaining assent. In order to ensure that the subject is able to make an informed voluntary decision, the study should be explained to the child/adolescent at a level that is understandable to the individual, taking into account age, maturity, psychological state, and English language proficiency. The process should provide an opportunity for the minor to express willingness or unwillingness to participate. Care must be taken to ensure that the process is free of coercion from parents and investigators. When the research context may compromise the voluntary nature of assent, particularly in vulnerable populations or in extremely sensitive situations, IRBs may consider the appointment of a participant advocate (an individual with no relationship to the research itself or the family of the participant, however, not necessarily a legal guardian) (Fisher et al. 1996).

It is critical to make sure that all children understand what is involved in the research study for which the investigator is trying to get assent. There are several ways to accomplish this. For example, the investigator may ask the subject to read the consent form aloud. Alternatively, to avoid embarrassment due to problems with reading or comprehension, a video or pictures may be used. The researcher can then discuss the content of the study with the minor, as well as ask questions about relevant content regarding the study. For laboratory studies or clinical procedures, actual demonstration, video or pictures of the procedures should be considered (e.g., simulation of the experience of being in an MRI machine). The capacity for decision-making may also be affected by substance use or abuse or comorbid disorders. Every effort must be made to develop procedures that document competence in understanding the study procedures and the risk/benefits of participating in the study. For example, procedures should be in place and staff should be qualified to determine that the potential participant is not under the influence of drugs or alcohol, under undue stress because of

withdrawal, or otherwise impaired in their cognitive or decision-making abilities while giving consent/assent.

Although typically consent/assent should be obtained before an individual is allowed to participate in a study, it is important to remember that an individual has the right to withdraw from a study at any time without penalty or loss of benefits. Therefore, procedures should also be in place that allow for the continued monitoring and ensuring of consent/assent to participate in research that is ongoing.

Waiving assent

The IRB can waive the assent requirement, as it can waive consent requirements, when it finds that (1) the research involves no more than minimal risk to the subjects; (2) the waiver will not adversely affect the rights and welfare of the subjects; (3) the research could not practicably be carried out without the waiver; and (4) whenever appropriate, the subjects will be provided with additional pertinent information after participation (45 CFR 46.116 (d)).

Box: Federal Risk and Benefit Categories for Pediatric Research

Prospect of Direct Benefit

Minimal Risk* Approvable by an institutional review board (IRB) provided:

- Parental permission†
- Child's assent‡

Minor Increase Over Minimal Risk Approvable by an IRB provided:

- Risks are "justified" by the anticipated benefit
- Risk-to-benefit profile is at least as favorable as the available alternatives
- Parental permission†
- Child's assent‡

More Than a Minor Increase Over Minimal Risk Approvable by an IRB provided:

- Risks are "justified" by the anticipated benefit
- Risk-to-benefit profile is at least as favorable as the available alternatives
- Parental permission†
- Child's assent‡

No Prospect of Direct Benefit

Minimal Risk* Approvable by an IRB provided:

- Parental permission†
- Child's assent‡

Minor Increase Over Minimal Risk Approvable by an IRB provided:

- Intervention is reasonably commensurate with subjects' actual or expected experience(s)
- Intervention is likely to yield generalizable knowledge about subjects' disorder or condition, which is of critical importance for the understanding or amelioration of the subjects' disorder or condition
- Parental permission
- Child's assent

More Than a Minor Increase Over Minimal Risk Not approvable by an IRB**

*Means "that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests" (** 46.102 [i]).

†Permission of 1 parent is sufficient for minimal risk and prospect of direct benefit research; permission of both parents is required in all other cases, if both are reasonably available. Parental permission may be waived if the IRB makes the findings under 45 CFR 46.116 (c) or (d) or judges that it is not a "reasonable requirement to protect the subjects" (**46.408 [c]).

‡May be waived if the IRB judges that the children are not capable of providing

assent, or the "research holds out a prospect of direct benefit that is important to the health or well-being of the children and is available only in the context of the research" (**46.408 [a]). Assent may also be waived under the provisions of 45 CFR 46.116 (c) and (d).

May be approved by the Secretary of the Department of Health and Human Services after consultation with a panel of experts and public review and comment, if the research satisfies the conditions of 45 CFR 46.404, 46.405, or 46.406 or (i) offers a "reasonable opportunity to further the understanding, prevention or alleviation of a serious problem affecting the health or welfare of children" (46.407), (ii) will be conducted in accordance with sound ethical principles, and (iii) adequate provisions are made for soliciting the assent of children and the permission of their parents or guardians.

Source: Modified from Shah et al. (2004); JAMA 291:476–482.

Specifically with respect to waiving the assent of minors, if the IRB determines that the capability of some or all of the children is so limited that they cannot reasonably be consulted, or that the interventions or procedures involved in the research hold out a prospect of direct benefit that is important to the health or well-being of the children and is available only in the context of the research, the assent of the children can be waived (45 CFR 46.408 (a)).

Consent from minors

Under applicable state law, emancipated minors are able to give independent consent. These minors have become emancipated for various reasons such as judicial decree, marriage or parenthood. They typically are financially independent and live away from home. The mature minor "is usually defined by state law as a minor that is near the age of maturity, displays sufficient understanding of medical procedures, and can be medically emancipated in the treatment of certain conditions, including venereal disease, pregnancy, and drug abuse" (American Academy of Pediatrics 1995). This legislation was intended to ensure that adolescents would not be deterred from seeking treatment (Levine 1995). Because Federal regulations define children as "persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research will be conducted," when state law allows minors of a specified age to consent to treatment for substance abuse, for example, these minors are no longer considered "children" for the purposes of research involving treatment for substance abuse. They, therefore, also have a legal right to consent to participate in a substance abuse treatment research protocol without the permission of a parent (Brody and Waldron 2000; English 1995). This exception only holds for research on a treatment for which they have a legal right to consent without parental permission. If the research protocol involves any procedure not related to this treatment, parental permission is required. Therefore, minors may not be able to consent for research procedures that are "add-ons" to the treatment. For example, a study examining whether a specific treatment works to prevent substance abuse, may require blood tests or behavioral assessments that are only used as research tools to determine the adequacy of the treatment. Procedures introduced solely for research purposes (1) must be considered separately from the treatment itself and (2) depending upon the relevant state law may not be allowed without parental permission. For more information on individual state laws regarding consent from minors, please see the Institute of Medicine Report, "The Ethical Conduct of Clinical Research Involving Children," Appendix B.

Obtaining permission from the emancipated or mature minor to inform parents about the study is preferred. However, in studies which involve minimal risk or in which benefits can be directly derived for the child, informing parents or informed permission from parents may not be necessary. Nonetheless, because participation in

treatment studies may involve different and potentially greater risks than standard treatments, parental permission is recommended.

Incarcerated children

In addition to the Federal regulations providing additional protections for children in research, if the research subjects to be studied are incarcerated minors, the research must comply with Federal regulations for research involving prisoners outlined in 45 CFR Part 46 Subpart C. According to Subpart C, "prisoner is defined as any individual involuntarily confined or detained in a penal institution. The term is intended to encompass individuals detained in other facilities by virtue of statutes or commitment procedures which provide alternatives to criminal prosecution or incarceration in a penal institution, and individuals detained pending arraignment, trial, or sentencing." Juveniles court-ordered to a residential treatment facility in lieu of incarceration are also considered "prisoners." This definition does not include individuals released from prison to a halfway house, those court-adjudicated to attend non-residential treatment programs or probationers and parolees. Any HHS-conducted or supported research involving prisoners must be certified by OHRP under Subpart C. In addition, if a research subject becomes incarcerated during the course of a study that was not previously approved in accordance with the requirements of Subpart C, appropriate certification must be provided to OHRP as soon as possible. For additional information, please see the OHRP Guidance on Research Involving Prisoners at http://www.hhs.gov/ohrp/humansubjects/guidance/prisoners.htm.

Wards of the state

Research may be conducted with children who are wards of the state or any other agency provided such research is (1) related to their status as wards or (2) conducted in schools, camps, hospitals, institutions, or similar settings in which the majority of children involved as subjects are not wards. If the research falls into one of these categories, an advocate must be appointed for each child who is a ward, in addition to any individual acting on behalf of the child as guardian or in loco parentis. This individual must have background and experience in acting in the best interests of the child and must not be associated in any way with the guardian organization (45 CFR 46.409).

C. Parental Permission

Because parents/guardians are responsible for protecting the children under their care, permission must be obtained from them to involve their children in a research protocol (except for emancipated minors as discussed above), even in the rare instance when a minor gives assent but a parent does not consent to permission. For research involving no more than minimal risk or involving greater than minimal risk but presenting the prospect of direct benefit to individual subjects, permission from one parent is sufficient. For other categories of research permissible under Subpart D, permission generally must be obtained from both parents, unless one parent is deceased, unknown, incompetent, or not reasonably available, or when only one parent has legal responsibility for the care and custody of the child.

Parental permission may be a particular challenge if the parent(s) experience problems with substance abuse, comorbid disorders or associated consequences of substance abuse including instability in their life circumstances. Therefore, parental permission must give adequate consideration of the mental and physical state of the individual in terms of their ability to fully understand the context of the informed consent document. In addition, the motives of the individual must also be considered. If there is a question about the parent or guardian's ability to give permission, procedures should be in place to evaluate the parents' ability to give or maintain permission, for example by a qualified independent third party. If the parent or guardian is deemed unable to give or maintain permission, the participant should be excluded from the study. Parental permission may be waived under the same conditions that allow waiver of consent as specified in 45 CFR 46.116 (c) and (d) and

discussed above in Section III B "Waiving assent."

D. Coercion and undue inducement

The issue of coercion is another area that may be especially salient among this population. Coercion comes in many forms. Fear of expulsion or incarceration cannot be used to coerce minors into participation. In situations where no treatment is readily available or treatment is cost prohibitive, and participating in a treatment study is the only access to treatment, parents may pressure their children to participate. Therefore, acquiring assent is critical for determining the willingness of a child to participate. Finally, the potential for remuneration may also be a factor leading to parental coercion for participation. Financial compensation to parents for their child's participation in a study should be commensurate with the requirements of the study (i.e., for effort, time and inconvenience of the research), as long as no "undue inducements" are offered to lure people into participating who would otherwise choose not to expose themselves to research risks and incentives are not included as a "benefit" in risk-benefit analyses. Children may also be compensated for participation, preferably with incentives other than money (e.g., vouchers). Offering evening or weekend hours and on-site childcare may assist parents who are concerned about lost wages to circumvent the need for monetary compensation.

E. Confidentiality

Investigators should be aware that once information from a drug-abusing youth is placed in the patient records, such records must be handled with extreme confidentiality, beyond those for other medical or research records. Investigators and IRBs should be aware that special federal requirements might apply to certain drug abuse records in research. Information about this may be found in the Code of Federal Regulations (CFR) under 42 CFR Part 2 (http://www.access.gpo.gov/nara/cfr/waisidx_02/42cfr2_02.html), "Confidentiality of Alcohol and Drug Abuse Patient Records" and, for covered entities, in the Privacy Rule under the Health Insurance Portability and Accountability Act (HIPAA—http://www.hhs.gov/ocr/hipaa/).

Confidentiality from parents

Because of the nature of the population and the subject matter to be studied, confidentiality is a particularly sensitive issue. Assurances of confidentiality must be given to the child/adolescent study participant as well as the parents or whatever other parties (e.g., schools) are involved. If children are excluded from a study because they do not satisfy eligibility requirements that include sensitive information about the child, investigators may not disclose these requirements to parents to protect the confidentiality of the child. During the consent/assent process, all parties should be informed that all information collected during the study remains confidential, and therefore not disclosed to the parent/guardian, including the use of illegal drugs, unless there is a risk of imminent danger to the child or to others, such as suicide or homicide. Given the nature of drug abuse research, there is an inherent danger to research participants who abuse drugs. However, determining whether or not this danger is imminent and, thereby, permitting disclosure of confidential information, is more complicated. There are a number of factors to consider, including whether or not the research participant is in treatment, which drug or drugs are being abused and by what route of administration, the presence of comorbid psychiatric conditions, and the age of the minor. Currently, there are insufficient data to provide specific guidelines on all situations that constitute imminent danger; therefore, we rely on the judgment of the investigator in consultation with their IRB. Moreover, the differentiation between use of drugs and imminence of serious harm resulting from the use of drugs should be clearly stated to the participants, individuals who have signed the informed consent, and relevant parties. The consent form should explicitly state what information may and may not be disclosed to the parent, child, or a third party such as a physician or mental health professional. The

consent form should also clearly specify that the release of information to the parents or other parties would only occur with the permission of the minor, except in the case of an adjudicated youth or the need to report withdrawal from the study to the referring agency. If relevant to the study, parents must be made aware that information on their own drug use or psychiatric history (or any other of their children who are minors) will be collected during the course of the study, but also kept confidential. Furthermore, both parties must be made clearly aware that any other mandated information such as domestic physical abuse, sexual abuse or neglect, and information on communicable sexually transmitted diseases will be reported to relevant agencies according to State law. Other than the mandated information, the final decision regarding the release of information to the parents resides with the investigator, even with signed authorization of the minor (Code of Federal Regulations, Title 42, 1995).

F. Community consultation

Given the sensitive nature of research on substance abuse involving children or adolescents, particularly that which does not present direct benefit to the child, it may be beneficial to establish an advisory board, including members of the community and advocacy groups, at the early stages of developing a research protocol (American Academy of Pediatrics 1995). Issues of race, ethnicity, socioeconomic status, and institutionalization that characterize a study population and issues of potential stigmatization should be considered at all stages of research design, development, and implementation.

IV. Specific Issues

The NACDA recommends that these specific issues be considered in the development and review of research studies involving children/adolescents.

A. Criteria for stopping a study

Prior to the initiation of a study, procedures for monitoring research subjects and clear criteria for when the study should be stopped due to increased levels of risk or decreased benefit should be established, whereas such criteria are typically established for pharmacotherapeutic studies, they should be established for behavioral research as well. One example in which the level of risk changed during a behavioral intervention study was when iatrogenic effects were observed during the course of group therapy with adolescents who engage in high-risk behaviors (Dishion et al. 2001). This study demonstrates the need to set specific criteria for increased risk/decreased benefit that will result in termination of any research study involving children. In addition, for pharmacological interventions, stopping rules should be in place for when there is incontrovertible evidence of benefit. This will necessitate an interim analysis with specific stopping rules for the Data Safety Monitoring Plan (DSMP) or Board (DSMB).

B. Competence of study staff

Sufficient expertise in child development, psychopathology, and ethical conduct of research with children and adolescents should be represented among the research staff.

C. Follow-up and referral

Because children and adolescents are considered to be a vulnerable population, careful monitoring and follow-up of this population is essential. In addition, if the adolescent or child is actively using substances, regardless of the type of study, a mechanism for referral to treatment should be established. Furthermore, for subjects in treatment studies that do not have access to treatment outside of the study, continued support should be identified and a concrete viable referral made prior to the end of the study.

D. Incidental clinical findings

During the course of a study, information may be obtained that is clinically meaningful, which is not directly addressed by the research protocol. For example, if an MRI is performed, information may be obtained indicating a previously unknown or undiagnosed medical problem. Similarly, in collecting information about a person's physical or psychological history, unexpected clinical findings may be made. It is important that the researcher state ahead of time how such information will be handled. To this end, and to the extent that the nature of such information can be predicted, the consent form should clearly stipulate what kind of incidental clinical findings will be provided and to whom. Should the circumstances or assumptions surrounding such provisions change significantly during the course of the study, the investigator should consider the introduction of appropriate changes in the consent form.

E. Studies involving administering drugs of abuse to youth

The NACDA Guidelines for Administration of Drugs in Human Subjects discuss issues that arise in research involving the administration of drugs with abuse/dependence liability, and identify issues to be considered in the development and review of research protocols involving drug administration to human subjects. The investigator is referred to this document on NIDA's website, http://www.drugabuse.gov/Funding/HSGuide.html.

Because adolescence is a period of heightened vulnerability to drug abuse and also a time of dramatic changes in brain and behavioral development, there may be compelling reasons to study how drugs of abuse specifically and/or uniquely affect young people. For studies that involve administration of drugs with abuse liability to youth, the decision to conduct a specific study will need to be made on an individual basis, carefully balancing the risks and benefits to participants. Some of the issues to consider in making this determination are the following: Are there sufficient safety data in adults or older adolescents (18–21 years) to warrant conducting the study in younger subjects? What is the most appropriate recruitment population? What is an acceptable level of risk for adolescents in a drug study? Will exposure under experimental conditions lead to future use? Will exposure send the wrong message about drug use? What are the risks associated with the administration of a drug to a youth, in light of ongoing developmental changes? For these sorts of questions, follow-up data collection would be critical. Who is responsible for the adolescents' safety, during the study, to and from the study, and after the study? What is the balance between confidentiality and investigator obligations within statutory regulations? What are the appropriate doses to use in youth? Consideration must be given to prior history of drug use, and family history of drug or mental health problems. Statutory regulations both among states and local authorities must be given careful attention.

Currently, NIDA does not fund any research in which drugs of abuse are administered to minors (including those who are current drug users). However, because of the importance of improving our understanding of how drugs of abuse affect the developing brain and behavior, future research questions may require such studies to be considered. Investigators should note that because the administration of drugs with abuse liability to children or adolescents may pose greater than a minor increase over minimal risk and no prospect of direct benefit, such protocols will most likely fall under 45 CFR 46.407. In order to provide better guidance and to emphasize the significance of considering specific issues during the development and review stage of such proposals, two examples of NIH-supported protocols, that have recently or are currently under review by DHHS according to 45 CFR 46.407, are described below.

1. "Alcohol, Sleep and Circadian Rhythms in Young Humans, Study 2—Effects of Evening Ingestion of Alcohol on Sleep, Circadian Phase, and Performance as a Function of

Parental History of Alcohol Abuse/Dependence" (funded by NIAAA). This protocol proposes to study the effects of a small or moderate evening dose of alcohol on sleep, waking performance, and circadian phase in a total of 64 adolescents (15–16 years of age) and young adults (21–22 years of age), and examine how the effects may differ between individuals who have a parent with a history of alcohol dependence and those who do not. The final recommendation by OHRP was that "HHS defer support for the proposed research involving the enrollment of 15- to 16-year old subjects…" because "adequate justification has not been provided…OHRP notes that ongoing IRB-approved studies under the grant will provide data relevant to both the safety of study subjects and the scientific rationale for involving 15- to 16-year old subjects. Upon completion of …ongoing research on adults…re-review of the proposed research would be warranted to consider extending the research to 15- to 16-year old subjects." The Acting Assistant Secretary for Health, HHS approved these recommendations.

2. Effects of single Dose of Dextroamphetamine in Attention Deficit Hyperactive Disorder: A Functional Magnetic Resonance Study" (funded by NIMH). This study proposes to investigate the pathophysiology of ADHD by imaging the brain response to amphetamine of children with ADHD compared to healthy children. The study is clearly of high significance since there is great public interest in the matter of stimulant treatment of ADHD. Because amphetamine would be administered to healthy children, this protocol was forwarded to, and is currently under review by, the OHRP under 45 CFR 46.407.

F. **Neuroimaging**

There are two categories of imaging techniques that can be used to investigate predispositions to and the effects of drugs of abuse; those that utilize magnetic fields such as functional magnetic resonance imaging (fMRI) and those that use ionizing radiation such as positron emission tomography (PET) and single photon emission computed tomography (SPECT).

There are three primary types of risks associated with PET and SPECT studies in minors, the stress associated with the procedure, the risks associated with arterial cannulation required by some studies to quantify radiotracer delivery, and the side effects associated with exposure to ionized radiation. Stress risks can be diminished by familiarizing children with the procedure by role rehearsal prior to the study or exposure to a simulated scanning device. The risks of arterial cannulation include mild-to-moderate pain, bruising at the puncture site, and spasm or clotting of the artery with a temporary decrease in blood flow. In rare instances blocking of the artery, poor healing, or infection at the catheter insertion site may occur. Permanent damage is extremely rare.

The risks of greatest concern have been those that are associated with radiation exposure such as potential carcinogenic effects or increased incidence of genetic mutations. According to the Federal Drug Administration Guidelines for use of radioactive drugs for research (21 CFR 361), "a single radiation dose for a research subject under 18 years of age to the whole body, active blood-forming organs, lens of the eye and gonads shall not exceed 0.3 rem and the annual total dose should not exceed 0.5 rem. For all other organs, a single dose cannot exceed 0.5 rem and the annual total dose cannot exceed 1.5 rem."

Exposure to radiation through neuroimaging research constitutes more than a minor increase over minimal risk. Therefore, this research would be approvable by an IRB only if study participants could directly benefit from the research. For all others, including healthy controls, this type of protocol would have to be reviewed under 407 provisions, thus could only be done if approved at the Department level. In addition, since the risks for radiation appear to reflect cumulative effects, cumulative exposure should

be considered when determining whether the risk is justified by the anticipated benefits, particularly if repeated use of neuroimaging with radioactive tracers is anticipated in a study.

Because of a lack of radiation exposure, fewer risks are encountered in fMRI studies. One concern with conducting fMRI (as well as PET/SPECT) studies in children, however, is the potential need for sedation to ensure that children remain still during the scans. Sedation presents significant risks due to potential complications such as respiratory distress. An alternative to sedation is to schedule scans for young children during times when children are naturally sleepy. This can also alleviate anxiety due to claustrophobia.

G. Genetics Studies

The primary ethical issues to be considered when undertaking genetics studies are the confidentiality of genetic information and the comprehension of the concepts of risk and probability associated with the identification of susceptibility genes. Genetic information must not be included as part of the medical record which may be subject to requests from insurance companies and employers. Investigators may alternatively determine that genetic information is maintained as part of a research record that is subject to the confidentiality guidelines discussed in the section entitled "Confidentiality" above. In this case, all information collected during the study remains confidential unless there is imminent danger to the child or to others. Since genetic susceptibility to substance abuse would not be considered to pose imminent danger, this information should also remain confidential.

Another concern in genetics studies is the risk of harm resulting from the lack of understanding of genetic findings. For example, subjects/parents may minimize the potential for prevention or behavioral change due to the misconception that genetic susceptibility to substance abuse necessarily means a subject will become addicted. Therefore, care must be taken to ensure that subjects and their parents understand that the genetics related to substance abuse may provide information about increased risk and information on better treatments for the disorder in the future but due to the multifactorial nature of substance abuse, is not deterministic for developing a substance abuse disorder. In many cases, genetic information is not released for this reason or because it is not yet clinically meaningful. If the investigator decides that it is inappropriate to release this information to the subject and/or parents, this should be made explicit during the consent process.

Provisions for removing samples from the study

Investigators must explain to potential subjects and their parents, during the informed consent process, about their options for removing samples from the study. If there is any reason why it may not be possible to remove samples in the future (e.g., DNA in repository that has been de-identified and distributed to other researchers), the timeline and reasons should be clearly elucidated.

Future use of DNA samples

DNA samples that are collected as part of a specific study may also be useful for future research not yet conceived. Consequently, participants may be given the opportunity to allow or deny the future use of their DNA samples for other purposes, or to ensure that personally identifying information is removed from their DNA samples before it is shared with other researchers or used for other purposes.

H. Survey Research

Research involving survey or interview procedures with children is not exempt from parental consent regulations (45 CFR 46.401). The No Child Left Behind Act of 2001 (Public Law 107-110 Section 1061) and the Protection of Pupil Rights Amendments (PPRA) delineate consent rules for surveys of students (34 CFR Part 98). Under the current law, if the US Department of Education (DOE) funds a study

and the research involves "protected information," then the PPRA afford parents the right to provide active consent. "Protected information" is defined as information on (1) political affiliations of student or student's parent; (2) mental or psychological problems of student or student's family; (3) sex behavior or attitudes; (4) illegal, anti-social, self-incriminating, or demeaning behavior; (5) critical appraisals of others with whom students have close family relationships; (6) legally recognized privileged or analogous relationships; (7) religious practices, affiliations or beliefs of student or student's parent; or (8) income. For studies that are funded by sources other than the US DOE (i.e., grants from the National Institutes of Health) and are administered by education institutions that receive funds from any US DOE program (i.e., public schools and some private schools), and that include protected information, parents have the right to inspect the surveys before they are administered and to opt the student out of the survey. The PPRA requires that individual schools adopt policies for consent requirements and IRBs must ensure that investigators use consent procedures that are in accordance with these local policies.

Waiver of parental permission

In some survey studies, specifically those that ask sensitive questions regarding illegal drug use and associated behaviors or environmental circumstances, it may be in the best interests of the study to acquire a waiver of parental permission by the IRB. Collecting this type of information in an anonymous manner may be crucial for detecting the prevalence of drug use and abuse and factors associated with increases or decreases in use. Therefore, field studies, which (1) anonymize the data; (2) pose minimal risk; (3) would be impractical to obtain parental permission (i.e., it would drastically lower response rate); and (4) would bias the results, may receive a waiver of parental permission by the IRB, under the same conditions that allow a waiver of consent as specified in 45 CFR 46.116 (c) and (d) and discussed above in Section III B "Waiving assent."

I. Treatment Studies

Placebo-controlled or untreated controlled studies

Researchers conducting studies of treatment versus placebo or untreated control groups must evaluate the risk/benefit ratio separately for the treated and untreated groups, i.e., those in the placebo group may not have a prospect of direct benefit whereas those in the treatment group do. Placebo or untreated observational control groups can be used in pediatric studies if their use does not expose children to unacceptable risks. For example, untreated control groups such as waitlist controls must be evaluated to determine the level of risk to which the participant is exposed by not receiving immediate care. In the case of placebo-controlled studies, the risks are acceptable when the potential risk for children in the placebo-control arm is equivalent to that for children receiving standard care or when the potential harms in the placebo-control arm are no more than a minor increase over minimal risk. In many such studies, treatment as usual is included to minimize undue risk in the placebo arm. Clear explanations of the purpose for and the consequences of being assigned to a placebo-control arm should be made during the informed consent process. Moreover, criteria for participant withdrawal, study discontinuation, and monitoring the status of the participant during the clinical trial should be clearly established as part of the research protocol prior to the study's initiation.

Behavioral interventions

Given the significant problem of substance abuse in youth and the unique considerations in treating this population, the development and testing of behavioral interventions specifically targeted for this population is critical. However, adequate attention must be paid to potential iatrogenic effects that may lead to increased risk for minor subjects so that such studies may be terminated if risk levels change. Therefore, as noted above, criteria for participant withdrawal, study

discontinuation, and monitoring the status of the participant should be clearly established prior to the study's initiation and described in the informed consent document.

Pharmacological interventions

New pharmacological interventions are emerging for the treatment of substance abuse, which may be useful in treating youth with substance use disorders. However, developmental differences may cause these drugs to behave differently in children than adults necessitating the study of these therapies in children. Pharmacotherapy studies in children should be performed after efficacy is established in adult Phase II studies. When data for adults are not available, particular care must be taken to justify proceeding with pediatric trials. Moreover, preclinical studies should be conducted using animals, during an equivalent developmental period, to assess potential developmental toxicity and efficacy before trials with children are started.

J. Pathophysiology Studies

Given that substance abuse is a developmental disease, beginning during childhood or adolescence, it is critical that studies of the pathophysiology of substance abuse, such as studies of basic biological mechanisms, are conducted with children. These studies may include those discussed above such as genetics and neuroimaging studies. However, because such studies offer no prospect of direct benefit, only those studies that present no more than a minor increase over minimal risk and are likely to yield generalizable knowledge about the disorder or condition may be approved. Because these studies often do not produce direct or immediate benefits to the participants, they may be the most sensitive and scrutinized types of research conducted in children. The recommendations noted above concerning multi-stakeholder advisory boards and/or consultation with members of the community prior to initiating such research may be particularly applicable to these kinds of scientific projects.

References

American Academy of Pediatrics, Committee on Drugs. (1995). Guidelines for ethical conduct of studies to evaluate drugs in pediatric populations (RE9503). *Pediatrics*, 95, 286–294.

American Lung Association. (2016). *E-cigarettes and lung health*. Chicago, IL: American Lung Association. Retrieved from: http://www.lung.org/stop-smoking/smoking-facts/e-cigarettes-and-lung-health.html

Befort, C., Lynch, R., James, R. L., Carroll, S. L., Nollen, N., & Davis, A. (2008). Perceived barriers to research participation among school administrators. *Journal of School Health, 78*(11), 581–586.

Brody, J. L., & Waldron, H. B. (2000). Ethical issues in research on the treatment of adolescent substance use disorders. *Addictive Behaviors, 25,* 217–228.

Brown, S. A., & Tapert, S. F. (2006). Adolescence and the trajectory of alcohol use: Basic to clinical studies. *Annals of the New York Academy of Sciences, 1021,* 234–244.

Casey, B. J., Getz, S., & Galvan, A. (2008). The adolescent brain. *Developmental Review, 28,* 62–77.

Center for Behavioral Health Statistics and Quality. (2015). *Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health* (HHS Publication No. SMA 15-4927, NSDUH Series H-50). Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from: http://www.samhsa.gov/

Center on Addiction and Substance Abuse (CASA) Columbia. (2011). *Adolescent substance use: America's #1 public health problem*. New York, NY: Columbia University Research Center. Retrieved from: http://www.casacolumbia.org/addiction-research/reports/adolescent-substance-use

Centers for Disease Control and Prevention (CDC). (2011). *School health programs: Improving the health of our nation's youth. At a glance 2011*. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion. Retrieved from: http://www.cdc.gov/chronicdisease/resources/publications/aag/dash.htm

Centers for Disease Control and Prevention (CDC). (2015a). *Youth risk behavior surveillance system*. Atlanta, GA: U.S. Centers for Disease Control and Prevention. Retrieved from: http://www.cdc.gov/healthyyouth/data/yrbs/index.htm

Centers for Disease Control and Prevention (CDC). (2015b). Tobacco use among middle and high school students-United States, 2011–2014. *Morbidity and Mortality Weekly Report*, 64(14), 381–385.

Chassin, L. (2008). Juvenile justice and substance use. *The Future of Children, 18*(2), 165–183.

Code of Federal Regulations. (2001). *Title 45 Part 46: Protection of human subjects*. http://www.nihtraining.com/ohsrsite/guidelines/45cfr46.html

Code of Federal Regulations. (2002a). *Title 42 CFR Part 2: Confidentiality of alcohol and drug abuse patient*

records. http://www.access.gpo.gov/nara/cfr/waisidx_02/42cfr2_02.html

Code of Federal Regulations. (2002b). *Title 45 Parts 160 and 164: The privacy rule under the Health Insurance Accountability and Portability Act (HIPAA)*. http://www.hhs.gov/ocr/hipaa

Code of Federal Regulations. (2003a). *Title 21 Part 361.1: Radioactive drugs for certain research uses*. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?FR=361.1

Code of Federal Regulations. (2003b). *Title 34 Part 98: Student rights in research, experimental programs and testing (Protection of Pupil Rights Amendments)*. http://www.access.gpo.gov/nara/cfr/waisidx_03/34cfr98_03.html

Colias, M. (2014, November 13). *After complaints, school district postpones behavioral surveys*. Rapid City, SD: Rapid City Journal. Accessed at: http://rapidcityjournal.com/news/local/after-complaints-school-district-postpones-behavioral-surveys/article_415b37e1-9b69-5894-8d97-e41e71044478.html

Cornelius, J., Kirisci, L., Reynolds, M., & Tarter, R. (2014). Does stress mediate the development of substance use disorders among youth transitioning to young adulthood? *American Journal of Drug and Alcohol Abuse, 40*(3), 225–229.

Dishion, T. J., Poulin, F., & Burraston, B. (2001). Peer group dynamics associated with iatrogenic effects in group interventions with high-risk young adolescents. *New Directions for Child and Adolescent Development, 91,* 79–92.

English, A. (1995). Guidelines for adolescent health research: Legal perspectives. *Journal of Adolescent Health, 17,* 277–286.

Fisher, C. B., Hoagwood, K., & Jensen, P. (1996). Casebook on ethical issues in research with children and adolescents with mental disorders. In K. Hoagwood, P. Jensen, & C. B. Fisher (Eds.), *Ethical issues in research with children and adolescents with mental disorders* (pp. 135–238). Hillsdale, NH: Erlbaum.

Green, J. (2014, July 28). White house responds to New York Times endorsing marijuana legalization. *The Weed Blog*.

Hingson, R. W., Heeren, T., & Winter, M. R. (2006). Age at drinking onset and alcohol dependence: Age at onset, duration, and severity. *Archives of Pediatric Medicine, 160*(7), 739–746.

Institute of Medicine. (2004). *The ethical conduct of clinical research involving children*. http://www.iom.edu/report.asp?id=19422

Johnston, L. D., O'Malley, P. M., Miech, R. A., Bachman, J. G., & Schulenberg, J. E. (2014). *Monitoring the future national results on drug use*. Ann Arbor: Institute for Social Research, The University of Michigan. Retrieved from: http://www.monitoringthefuture.org/pubs/inthemonographs/mtf-overview2013.pdf

Kann, L., Kinchen, S., Shanklin, S. L., Flint, K. H., Hawkins, J., Harris, W. A., et al. (2014). Youth risk behavior surveillance-United States, 2013. *Morbidity and Mortality Weekly Report, 63*(4), 13–24.

Lavrakas, P. J. (Ed.). (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.

Levine, R. J. (1995). Adolescents as research subjects without permission of their parents or guardians: Ethical considerations. *Journal of Adolescent Health, 17,* 287–297.

Macquarie University. (2006). *MQ HREC use of passive 'opt-out' consent in research involving children*. Sydney, AUS: Macquarie University. Available from: research.mq.edu.au/…/MQ_HREC_passive_consent_in_child_research_policy-4.doc

Marshal, M. P., Burton, C. M., Chisolm, D. J., Sucato, G. S., & Friedman, M. S. (2013). Cross-sectional evidence for a stress-negative affect pathway to substance use among sexual minority girls. *Clinical Translational Science, 6*(4), 321–322.

McConnell, M. M., Memetovic, J., & Richardson, C. G. (2014). Coping style and substance use intention and behavioral patterns in a cohort of BC adolescents. *Addictive Behaviors, 39*(10), 1394–1397.

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay and bisexual populations: conceptual issues and research evidence. *Psychological Bulletin, 129*(5), 674–697.

National Addiction & HIV Data Archive Program (NAHDAP). (2012). *About NAHDAP*. Retrieved from: http://www.icpsr.umich.edu/icpsrweb/NAHDAP/

National Advisory Council on Drug Abuse. (2000). *Guidelines for administration of drugs to human subjects*. http://www.drugabuse.gov/Funding/HSGuide.html

National Alliance to End Homelessness. (2015a). *LGBTQ youth*. Washington, DC: National Alliance to End Homelessness. Retrieved from: http://www.endhomelessness.org/pages/lgbtq-youth

National Alliance to End Homelessness. (2015b). *Youth*. Washington, DC: National Alliance to End Homelessness. Retrieved from: http://www.endhomelessness.org/pages/youth

National Institute on Alcohol Abuse and Alcoholism. (2006). Early drinking linked to higher lifetime alcoholism risk. *NIH News*, July 3. Retrieved from: http://www.nih.gov/news/pr/jul2006/niaaa-03.htm

National Institute on Drug Abuse (NIDA). (2006, July). *Principles of drug abuse treatment for criminal justice populations: A research-based guide*. NIH Publication No. 06-5316. Bethesda, MD: NIDA.

National Institute on Drug Abuse (NIDA). (2012). *NACDA guidelines for substance abuse research involving children and adolescents*. Bethesda, MD: NIDA. Retrieved from: http://www.drugabuse.gov/sites/default/files/pdf/nacdaguidelines.pdf

National Institute on Drug Abuse (NIDA). (2014, December 16). *Teen prescription opioid abuse, cigarette, and alcohol use trends down.* Bethesda, MD: NIDA. Retrieved from: http://www.drugabuse.gov/news-events/news-releases/2014/12/teen-

prescription-opioid-abuse-cigarette-alcohol-use-trends-down

National Resource Center on Domestic Violence. (2013). *Runaway and homeless youth and relationship violence toolkit*. Washington, DC: NRCDV. Retrieved from: http://www.nrcdv.org/rhydvtoolkit/each-field/homeless-youth/define.html

O'Donnell, L. N., Duran, R. H., San Doval, A., Breslin, M. J., Juhn, G. M., & Stueve, A. (1997). Obtaining written parent permission for school-based health surveys of urban young adolescents. *Journal of Adolescent Health, 21*(6), 376–383.

Office of Adolescent Health. (2014). *Substance abuse: Risk and protective factors*. Rockville, MD: U.S. Department of Health & Human Services, Office of Adolescent Health. Retrieved from: (http://www.hhs.gov/ash/oah/adolescent-health-topics/substance-abuse/tobacco/risk-and-protective-factors.html

Office of Human Research Protections. (2001). Protections for children in research: A report to congress in accordance with Section 1003 of P.L. 106-310, Children's Health Act of 2000.

Pokorny, S. B. (2006). *Active vs. passive parental consent*. Durham, NC: 3C Institute. Available at: http://www.4researchers.org/articles/146

Public Law 107-110 Section 1061. (2001). The No Child Left Behind Act. http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Range, L., Embry, T., & MacLeod, T. (2001). Active and passive consent: A comparison of actual research with children. *Ethical Human Science Services, 3*(1), 23–31.

Rew, L., Taylor-Seehafer, M., & Thomas, N. (2000). Without parental consent: conducting research with homeless adolescents. *Journal for Specialists in Pediatric Nursing, 5*(3), 131–138.

Secor-Turner, M., Sieving, R., Widome, R., Plowman, S., & Vanden, B. E. (2010). Active parent consent for health surveys with urban middle school students: processes and outcomes. *Journal of School Health, 80*(2), 108–110.

Shah, S., Whittle, A., Wilfond, B., Gensler, G., & Wendler, D. (2004). How do institutional review boards apply the federal risk and benefit standards for pediatric research? *Journal of the American Medical Association, 29,* 476–482.

Society for Research on Child Development. (1991). Ethical standards for Research with Children, Society for Research on Child Development. http://www.srcd.org/about.html

Squeglia, L. M., Jacobus, J., & Tapert, S. F. (2009). The influence of substance use on adolescent brain development. *Clinical EEG Neuroscience, 40*(1), 31–38.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2014). *Results from the 2013 national survey on drug use and health: Summary of national findings*. Rockville, MD: SAMHSA. Retrieved from: http://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Web/NSDUHresults2013.pdf

Tavolacci, M. P., Ladner, J., Grigioni, S., Richard, L., Villet, H., & Dechelotte, P. (2013). Prevalence and association of perceived stress, substance use and behavioral additions: a cross-sectional study among university students in France, 2009-2011. *BMC Public Health, 13,* 7324.

Trudeau, M. (2009, February 5). More students turning illegally to "smart drugs". *NPR News*. Retrieved from: http://www.npr.org/templates/story/story.php?storyId=100254163

UNC Carolina Population Center. (2015). *Add health*. The National Longitudinal Study of Adolescent to Adult Health. Retrieved from: http://www.cpc.unc.edu/projects/addhealth

US Department of Health and Human Services. (2003). *OHRP guidance on the involvement of prisoners in research*. Washington, DC: DHHS. Retrieved from: http://archive.hhs.gov/ohrp/humansubjects/guidance/prisoner.htm

US Department of Health and Human Services. (2014). *Healthy people 2020. 2020 topics and objectives. Substance abuse*. Washington, DC: DHHS. Retrieved from: https://www.healthypeople.gov/2020/topics-objectives/topic/substance-abuse

Witt, E. D. (2010). Research on alcohol and adolescent brain development: Opportunities and future directions. *Alcohol, 44*(1), 119–124.

Wolfenden, L., Kypri, K., Freund, M., & Hodder, R. (2009). Obtaining active parental consent for school-based research: A guide for researchers. *Australia New Zealand Journal of Public Health, 33*(3), 270–275.

LGBT Populations and Substance Abuse Research: An Overview

16

Dianne L. Kerr and Willie H. Oglesby

16.1 Introduction

Much previous research has demonstrated that lesbian, gay, bisexual, and transgender (LGBT) individuals report higher levels of substance abuse and related consequences than their heterosexual counterparts. However, due to the many methodological issues discussed in this chapter, the actual prevalence of substance abuse among LGBT people and the socio-demographic determinants and causal relationships with substance abuse behaviors in this population are difficult to discern. In this chapter, we review the current state of the literature on substance abuse among LGBT people and focus on several key methodological challenges that scholars face when researching substance abuse with this population.

Although the LGBT population is usually studied as a group, it is important to note that not all LGBT people are the same. They have different demographic characteristics, risk factors, and health outcomes. LGBT communities also contain subpopulations that vary by race/ethnicity, socioeconomic status, age, geography, culture, and other factors: just like their heterosexual counterparts. In some LGBT populations, the risk of substance abuse may be higher due to stigma, internalized homophobia, social norms, access, and other factors. For other LGBT populations, the risk may be lower due to increased resilience resulting from previous experience with prejudice, stigma, and other negative experiences that have increased their coping mechanisms, adaptability, and vigilance. This is similar to heterosexual populations—some experience higher rates of substance abuse and some have greater resilience.

16.2 Self-reported Substance Abuse

Due to the LGBT population's relatively small size, most of the previous research published on LGBT people was not drawn from population-based samples. Many of these prior studies used convenience or purposive sampling techniques, which limit the generalizability of their findings. Although most of what is known about LGBT people arises from these relatively weaker research designs, the information gleaned from these studies can be helpful in understanding the risk and protective factors, health-compromising behaviors, and health outcomes of LGBT people. For instance, previous research has demonstrated that LGBT populations have the highest rates of alcohol use (Cochran et al. 2000; Lewis et al. 1982; McKirnan and Peterson

D.L. Kerr (✉)
College of Education Health, and Human Services, Kent State University, Kent, OH, USA
e-mail: dkerr@kent.edu

W.H. Oglesby
College of Population Health, Thomas Jefferson University, Philadelphia, PA, USA
e-mail: Billy.Oglesby@jefferson.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_16

1989; Ziebold and Mongeon 1982), tobacco use (Diamant et al. 2000; Lee et al. 2009), and marijuana use (McKirnan and Peterson 1989; Drabble and Trocki 2005). Research has also shown that illicit drug use is higher among LGBT persons (McKirnan and Peterson 1989), with some variation noted across LGBT subgroups (Hughes and Eliason 2002; Hughes 2005).[1]

More current research uses population-based samples to examine LGBT health, including substance abuse.[2] Compared to early studies, much of this newer work employed more sophisticated research methodologies, including, but not limited to, improved sampling, measurement, and more in-depth investigations of multiple dimensions of sexual orientation and their relationship to substance use and abuse (Green and Feinstein 2012). The assessment of multiple dimensions of sexual orientation, in particular, was considered a major methodical advancement, as researchers could now explore in more nuanced detail outcomes associated with at least three components—sexual attraction, sexual behavior, and sexual identity, leading to a more accurate picture of substance abuse within this population (Green and Feinstein 2012).[3]

This picture only partially confirmed earlier findings with regard to substance abuse. For instance, with regard to alcohol abuse disorders and related problems, higher levels were detected, but largely among lesbian and bisexual women (Drabble et al. 2005; Green and Feinstein 2012). With regard to drug abuse, homosexually and bisexually experienced men and women are more likely to be at risk of addiction across all drug classes (Cochran et al. 2004; Green and Feinstein 2012; Stall et al. 2001).[4] Additionally, they reported higher levels of use of novel and/or club drugs (Corliss et al. 2009; Measham et al. 2011; Parsons et al. 2006; Rukus et al. 2016).[5] This picture is explored in greater detail in the following sections.

16.2.1 Substance Abuse and Consequences Among Lesbians and Bisexual Women

In 2013, the National Health Interview Study (NHIS) included sexual orientation questions for the first time. Sexual orientation results determined 27.2% of women who identified as gay or lesbian and 29.4% of women who identified as bisexual were current smokers, compared with 16.9% of heterosexual "straight" women. The NHIS results also found higher percentages of women aged 18–64 who identified as gay or lesbian (27.7%) or bisexual (34.9%) had five or more drinks in one day at least once in the past year compared to 17.2% of heterosexual women (Ward et al. 2014). Drabble and Trocki (2005)

[1] A full review of early studies of substance abuse among LGBT populations is beyond the scope of this chapter. See Bux (1996) and Green and Feinstein (2012) for more detailed methodological critiques of early studies.

[2] This change was accelerated after a formal work group was established in 2010 to examine scientific literature on LGBT health, which resulted in the Institute of Medicine Report on the Health of LGBT People (2011). This report revealed many health disparities of LGBT people. In 2011, Health and Human Services Secretary Kathleen Sebelius announced an HHS plan to enhance data collection on LGBT individuals in order to address health disparities. This led to the inclusion of sexual orientation questions on some national population-based surveys that will be explored later in this chapter. To date, few surveys include gender orientation questions including the Behavioral Risk Factor Surveillance System (BRFSS) and the Youth Risk Behavior Surveillance System (YRBS), although not all states or school districts, respectively, use these measures and more consistent use is needed (The GenIUSS Group 2014).

[3] For a detailed discussion of the three dimensions of sexual orientation and their association with psychological outcomes, see Savin-Williams (2006).

[4] In at least one longitudinal study, age was an important modifier of risk, with differences between sexual minority and heterosexual subjects greatest in adolescence (Corliss et al. 2009).

[5] LGBT people have been described as "early adopters" of new drug trends, particularly club drugs and novel psychoactive substances (Measham et al. 2011). People who use club drugs often more likely report poly-drug use, or the use of more than one substance (Halkitis et al. 2007; Halkitis and Palamar 2008). Early adoption and use of novel drugs have been linked to cultural definitions which destigmatize and reinforce use in the LGBT community (Rukus et al. 2016).

also found alcohol consumption to be higher among lesbian, bisexual, and heterosexual women reporting same-sex partners compared to heterosexual women who reported opposite-sex partners only. Lesbian and bisexual women also have greater odds of reporting alcohol-related consequences and alcohol dependence than heterosexual women (Drabble and Trocki 2005).

The college years are a time of life when alcohol abuse typically occurs. Although most previous studies have determined that lesbian and bisexual women consume more alcohol, a few studies of college women have found that lesbians are not significantly different than their heterosexual counterparts in their alcohol use, although the consequences they experience due to their drinking may be far greater (Kerr et al. 2014; McCabe et al. 2003). In one study, the consequences for college women that self-identified as lesbians included over 4 times greater odds of seriously considering suicide and over 3 times the odds of having sex without getting consent (Kerr et al. 2014). In another study of college women, consequences included more likelihood of driving under the influence, having unplanned sex, having suicidal thoughts, and sexually harassing someone after drinking (McCabe et al. 2003).

Drabble et al. (2005) found lesbian and bisexual adult women have greater odds of reporting alcohol-related consequences and alcohol dependence. Risky alcohol use and related problems are even higher for bisexual women as compared to lesbians or heterosexual women (Hughes et al. 2010; Drabble et al. 2005). In addition, bisexual women experience a higher number of lifetime sexual partners and an increased vulnerability to sexual victimization (Hequembourg et al. 2013) although the relationship between substance abuse and sexual victimization among lesbian and bisexual women needs further study (Eaton et al. 2008; Hequembourg et al. 2013).

Lesbian and bisexual women who seek treatment for alcohol abuse report that they are less satisfied with the treatment experience than heterosexual women (Drabble and Trocki 2005). Other substance abuse issues of lesbians and bisexual women include increased rates of smoking and marijuana use as compared to heterosexual women (Trocki et al. 2009). Illicit drug use appears to be higher among bisexual women than lesbians. Cochran et al. (2004) found that lesbian and bisexual women (defined as women with any female partners) were more likely to use marijuana and analgesics in the past 30 days than heterosexual women (defined as women with only male partners). Previous research also shows the increased use of club drugs such as meth and ecstasy among bisexual women as compared to heterosexual women (Parsons et al. 2006; Kelly and Parsons 2007; Boyd et al. 2003). Parsons et al. (2006) found lesbian and bisexual women more likely to use ketamine, LSD, and meth than heterosexual women. Scheer et al. (2002) found bisexual women more likely to report past and recent use of injection drugs including heroin, cocaine, and speed. In fact, several studies have now determined that bisexual women have greater ATOD use than any other sexual orientation group (Cochran et al. 2004; Corliss et al. 2009; McCabe et al. 2005; Scheer et al. 2002).

16.2.2 Gay and Bisexual Men

Recent results of the National Health Interview Survey (2013) indicated that a higher percentage of adults who identify as gay or lesbian (27.2%) or bisexual (29.5%) were current cigarette smokers as compared to heterosexuals (19.6%). However, when prevalence of current cigarette smoking by sexual orientation was examined for men, no significant differences were found among men aged 18–64. For alcohol consumption among men, a higher percentage of those who identified as bisexual (56.3%) had five or more drinks in one day at least once in the past year as compared to 35.1% of those who identified as "straight" (Ward et al. 2014).

In the 1990s and 2000s, the use of MDMA (ecstasy), Ketamine ("special K"), and other "club drugs" were sharply experienced in urban LGBT communities. It was also prevalent in the "rave communities" of mostly younger

heterosexual adolescents in urban areas. By the 2000s, the use of crystal methamphetamine (or "crystal meth") grew to epidemic proportions among the LGBT community and quickly spread to other groups (Hirshfield et al. 2006), including rural areas. In the 2010s, the use of crystal meth remains a public health crisis for many populations, as does the dramatic increase in opioid consumption.

Substance abuse among gay men has also been linked to violence in romantic relationships, where substance abuse increases physical violence, which in turn increases substance abuse in a cyclical manner (Andrasik et al. 2013). Among those in noncommitted relationships, gay men who reported the use of poppers, crystal meth, cocaine, marijuana, Viagra, or alcohol use before sexual encounters were significantly more likely to report unprotected anal intercourse, which dramatically increases the risk for sexually transmitted diseases, including HIV (Hirshfield et al. 2006). In addition, the use of "party drugs" was most often associated with sexual risk taking, although substance use before or during sex was not associated with risk with HIV-negative partners, but was associated with risk with HIV-positive and unknown serostatus partners (Purcell et al. 2005). In a study of gay and bisexual men who recently became HIV-positive, high rates of unprotected anal intercourse and recreational drug use took place at the high-risk event (Volk et al. 2006). In another study, men who have sex with men (MSM) who reported using methamphetamines were more likely to report higher education, health insurance coverage, inconsistent condom use during anal intercourse, a history of STD's, positive HIV status, and use of medication designed to treat erectile dysfunction (Rhodes et al. 2007). The health disparities of young gay and bisexual men such as substance abuse, partner violence and victimization, and HIV infection have led HIV prevention researchers to call for more integrated prevention programs that provide community-level interventions, gay-specific sexuality and HIV prevention education and substance abuse education, and access to mentors, goal setting, and future planning (Lyons et al. 2013).

16.2.3 Transgender Individuals

There is a dearth of research on transgender people overall, yet studies that have been conducted indicate they are at risk for a variety of health-compromising behaviors. In 1999 the San Francisco Public Health Transgender Risk Behavior Study found the majority of male-to-female (MTF) transgender individuals had a history of non-injection drug use with rates of 90% for marijuana, 66% for cocaine, 57% for speed, 52% for LSD, 50% for poppers, 48% for crack, and 24% for heroin. Further, 16 and 23% of MTF respondents had been in alcohol treatment or drug treatment programs, respectively (San Francisco Department of Public Health 1999).

One of the largest studies of transgender people conducted to date was the National Transgender Discrimination Survey Report on Health and Health Care, which was conducted as a joint effort of the National Center for Transgender Equality and the National Gay and Lesbian Task Force (Grant et al. 2010). This survey research study included over 7000 respondents. Researchers found over 25% of transgender respondents indicated they misused drugs or alcohol at some point to cope with the discrimination they faced. Eight percent of the sample said they were currently using drugs or alcohol to cope and 18% said they had done so in the past but were not currently doing so. Thirty percent of the sample reported smoking daily or occasionally. Alcohol and drug use decreased with age in the transgender group, similar to studies of the general population (Grant et al. 2010).

16.2.4 LGBT Youth

For many LGBT youth, substance use is initiated at a young age. Studies of Youth Risk Behavior Survey (YRBS) data comparing LGBT high school youth to heterosexual youth in illicit drug use demonstrate higher prevalence among sexual minority youth, with bisexual youth having the highest prevalence of all. The difference between lesbians and heterosexual

high school girls was not as great as those between gay and heterosexual high school boys (Newcomb et al. 2014).

Substance abuse is also associated with many health risk behaviors and problems among LGBT adolescents such as depression, suicide ideation and attempts, violence, victimization, and higher rates of unprotected sexual intercourse (Kerr et al. 2014). For this reason, youth interventions are needed both at the institutional and individual levels (Newcomb et al. 2014). A number of reasons have been posited for early initiation of substance abuse among LGBT populations, including using them as a short-term strategy to numb or escape the emotional or physical effects of abuse (Andrasik et al. 2013).

16.3 Minority Stress, Stigma, and Substance Abuse

Researchers have also linked stigma and discrimination to mental health issues and substance abuse among LGBT populations (Bux 1996; Hughes and Eliason 2002). Ilan Meyer's work on minority stress indicates LGBT individuals have more chronic stress due to stigmatization. He measured stigma using three constructs: (1) internalized homophobia, (2) stigma related to expectations of rejection and discrimination, and (3) actual experiences of discrimination and violence (Meyer 2003). Molina and Ramirez-Valles (2013) measured these three dimensions of stigma (internalized, perceived, and enacted) and found all of the stigma dimensions were associated with drug use, and all except perceived stigma were associated with alcohol use. McCabe et al. (2010) linked past year substance abuse disorders to discrimination. LGBT adults who experienced all three types of discrimination studied (sexual orientation, race, and gender) had nearly 4 times greater odds of past year substance use disorders. The researchers concluded that health professionals should consider the role of discrimination in the development of substance use disorders and their treatment. Much current research on LGBT health disparities uses Minority Stress Theory (Meyer 2003) as a theoretical underpinning.[6]

16.4 Methodological Problems in LGBT Research

Research with LGBT populations presents many challenges. These challenges involve varied definitions of sexual orientation, inaccurate assessments of the size of the LGBT population, inadequate sampling, a lack of research among certain groups in the LGBT spectrum, and combining groups of the LGBT spectrum for analysis purposes.

16.4.1 Varied Definitions of Sexual Orientation

Defining sexual orientation has been an ongoing challenge in substance use research, and researchers use various definitions. Three components of sexual orientation are presented in the literature. These components include sexual/romantic attraction, sexual behavior, and self-identification (Hughes and Eliason 2002). Often survey research questionnaires do not include all three of these components. Researchers frequently use self-identification or sexual behavior as measures or combine them. There appears to be less use of the sexual/romantic attraction component to identify these groups. Any of these components used alone may be problematic. For example, some investigators use the behavioral definitions and combine lesbian and bisexual women into one group of "women who have sex with women (WSW)." Similarly, gay and bisexual men are combined into one group of "men who have sex with men (MSM)."

[6]While social stress and discrimination constitute the most widely accepted risk factors for substance abuse in the LGBT population, other explanatory factors have also been attributed as contributing to high rates of substance abuse. This includes levels of affiliation with gay and lesbian subcultures and peer influences which promote use (Green and Feinstein 2012). For a review of the role of social networks, see McCrady (2004).

One definition of this behavioral classification is "any mutually voluntary activity with another person that involves genital contact and sexual excitement or arousal, that is feeling really turned on, even if intercourse or orgasm did not occur" (Laumann et al. 1994). This behavioral definition may result in an overrepresentation of gays and bisexuals as using this classification may cause one to self-identify as gay or bisexual as a result of a one-time same-sex encounter, when this was merely sexual experimentation, a common experience among young people.

Another problem with research using the men who have sex with men or women who have sex with women categorizations is that they combine gays and bisexuals into one classification for analysis when these groups are often very different from each other. The most recent Institute of Medicine (IOM) report advises against this practice, recommending that groups of the LGBT spectrum be separated out for analysis rather than combined (IOM 2011). In attempts to further delineate classifications of lesbians, investigators sometimes use terms such as "lifetime lesbian" in order to distinguish between lesbians that have identified as such throughout their lives versus "adult lesbians" who identified as lesbians later in life. This classification was used in a study of the Women's Health Initiative data, for women from 50 to 79 years old (Valanis et al. 2000). Investigators used the term "adult lesbians" to classify those who identified as lesbians after 45 years of age.

In terms of the sexual identity classification, some researchers have used a more varied description for sexual identity than the traditional three-category classification (heterosexual, bisexual, or gay/lesbian) or a four-category definition that adds "unsure" to the three-category classification. A five-category description of sexual orientation (only heterosexual, mostly heterosexual, bisexual, mostly lesbian/gay, only lesbian/gay) has led to different results on some measures of substance use when compared to the traditional three-category classification (McCabe et al. 2012). There is also a 6-option categorization: (1) gay or lesbian; (2) bisexual, but mostly gay or lesbian; (3) bisexual, equally gay/lesbian; (4) bisexual, but mostly heterosexual; (5) heterosexual; and (6) uncertain, do not know for sure (D'Augelli et al. 2001) and a classification that includes additional options such as "curious", "questioning", and "unlabeled" (Thompson and Morgan 2008). These many classifications complicate research with LGBT populations.

Sexual/romantic attraction is the third component of sexual orientation. It is defined as "attraction toward one sex or the desire to have sexual relations or to be in a primary loving, sexual relationship with one or both sexes" (Savin-Williams 2006). Survey questions related to this concept typically ask whether the individual has had a romantic attraction to someone of the same sex.

More perplexing for researchers is the fact that prevalence of homosexuality reported by respondents is often different depending on which definition is used. For example, those with same-sex behavior may not self-identify as homosexual. This phenomenon, while not exclusive to African American men, is described in that population as being "on the down low." The term "on the down low" in HIV circles is often applied to African American men who identify as heterosexual and have a female partner, while having sex with other men in secret. While these men are often blamed for expanding the HIV epidemic, this argument ignores other possible contributing factors such as the high prevalence of sexually transmitted infections (STIs) (AVERT 2014) or the number of African American men who may be infected with HIV while in prison. Malebranche (2011) suggests "what influences sexual identification among Black MSM may be a complex mix of racial and racist life experiences, gender norms, religious beliefs and masculine socialization."

Individuals often respond inconsistently to the three components of sexual orientation in survey research and frequently change their responses over time. Different answers to the three component questions may cause one to be classified as homosexual in one study but not in another. Savin-Williams stated "until conceptually well-positioned and psychometrically sound and

tested definitions are used, it is unlikely that research can possibly or reliably identify the prevalence, causes and consequences of homosexuality" (Savin-Williams 2006). Thus, varied definitions of sexual orientation continue to present a major obstacle in LGBT research and do not allow for comparisons across studies. Only a few studies have included all three dimensions of sexual orientation (McCabe et al. 2005, 2009; Goldberg et al. 2013).[7]

Question wording is also important to determine sexual or gender orientation. For example, if transgender is included as a gender selection on a questionnaire (e.g., male, female, and transgender), a two-step method is needed. A question on current gender identity should be followed up with a question on sex assigned at birth (Conron et al. 2014). If the two steps (questions) are not included it may be impossible to determine the current gender identification. For example, the American College Health Association's National College Health Assessment (ACHA-NCHA-II) appropriately included "transgender" as a choice in the gender demographic but failed to include a follow-up question, making it impossible to determine whether respondents were a male-to-female or female-to-male transgender individual making it impossible to analyze differences between the transgender groups. Therefore, a researcher studying transgender college students with this data set should note this as a major limitation. The American College Health Association recently remediated this problem with a revision of the ACHA-NCHA instrument.

As a model for future research, The Fenway Institute (2013a, b) recommends a two-step gender identity question to determine both the current gender identification and the sex assigned at birth for use in the Electronic Medical Record (EMR). Table 16.1 shows a two-step version for the EMR approved by the World Professional Association for Transgender Health (WPATH) and the Centers for Disease Control and Prevention (Deutsch et al. 2013).

The Fenway Institute utilized very similar two-step questions in a recent study in four community health clinics. The 2-step question performed well in the study with nearly zero missing data and twice the response rate of a one-question format (The Fenway Institute 2013a, b). Another organization interested in how to ask survey questions on sexual orientation and gender identity is the Williams Institute. They formed a group of experts known as the Gender Identity in U.S. Surveillance (GenIUSS) group, to determine current practices to identify transgender and other gender minority respondents in population research. The GenIUSS group developed a publication addressing how to ask gender identity questions on population-based surveys in 2014.[8] Examples of national data sets that include sexual orientation and/or gender identity questions are included in the Appendix.

16.4.2 Varying Estimates of the Size of the LGBT Population

Research has also been directed at estimating the size of the LGBT population for the purposes of calculating prevalence and incidence rates in the United States. Laumann et al. (1994) estimated that the percent of the population who were MSM was 1% in rural areas, 4% in suburban areas, and 9% in urban areas. Findings from the National Survey of Family Growth found that an estimated 6.0% of randomly sampled males aged 15–44 in the US reported ever having sexual contact with another male (Mosher et al. 2005). Lieb et al. (2009) created, constructed, and compared different population models and found MSM prevalence rates between 3.7 and 17.0% in

[7]McCabe et al. (2005) were the first to provide a detailed analysis of substance use patterns by dimensions of sexual orientation. This study found that women who identified as mostly heterosexual, bisexual, or who were attracted to/had sex with both men and women reported higher rates of substance use. For men, mostly heterosexual identity was associated with higher drug use compared to only heterosexual identity.

[8]The GenIUSS publication is available online at http://williamsinstitute.law.ucla.edu/wp-content/uploads/geniuss-report-sep-2014.pdf.

Table 16.1 Two-step method for the collection of sex and gender identity information

Current gender identity	Sex assigned at birth
Male	Male
Female	Female
Transmale/transman/FTM	Other
Transfemale/transwoman/MTF	
Genderqueer/gender-non-conforming	
Different identities: please state_____	

FTM female to male (i.e., female assigned at birth, male-spectrum identity); *MTF* male to female (i.e., male assigned at birth, female-spectrum identity) (Deutsch et al. 2013)

Southern US states and the District of Columbia. Most recently, findings from the National Health Interview Survey (NHIS) found 1.6% of the respondents identified as gay or lesbian, 0.7% identified as bisexual, and 1.1% of adults identified as "something else" stated "I don't know the answer" or refused to provide an answer (Ward et al. 2014). It is important to note, however, that the NHIS is administered by face-to-face or phone interviews, which may cause participants to be reluctant to reveal their sexual orientation. For topics that may be too sensitive to discuss in face-to-face interviews, audio-equipped computers may maximize responses, although NHIS administrators chose not to use this technology for the sexual orientation questions during this first administration of the questionnaire that included these questions.

16.4.3 Sampling Issues

Lack of consensus on the size of the LGBT population has direct implications for sampling. The LGBT research literature often uses small samples of discrete groups of LGBT individuals (e.g., members of a campus LGBT Pride Group, an LGBT community center, or those attending a gay pride event) which may or may not be representative of larger groups of LGBT individuals in the given setting or community. The research literature is rife with studies using these small nonrepresentative samples (Green and Feinstein 2012). In the past, access to "out" LGBT individuals was limited to venues where they may be present in greater numbers. Before the Stonewall Rebellion, LGBT individuals lived in fear of harassment and violence. Bars were one of few meeting places, and even in these settings LGBT individuals were subject to police harassment and arrest. However, after Stonewall, gay bars were used to conduct research on substance use and misuse (Fifield et al. 1977). Certainly, there are limitations to conducting substance use research in a bar setting. As a consequence of the bars being one of the few venues to capture LGBT populations, much of the early research found increased tobacco and alcohol use among LGBT individuals. The validity of such research certainly comes into question, although more recent studies with representative populations continue to show LGBT groups have increased ATOD use. Additionally, snowball sampling techniques were also used frequently in the past. In this type of sampling, LGBT individuals are given surveys to complete and asked to provide them to LGBT friends. These convenience samples are not considered representative of the LGBT population, but were one of the few ways to collect data on this often-hidden population.

While venue-based and other convenience sampling techniques continue to be used, increasingly researchers have begun to utilize national- and state-based samples.[9] Today, larger national health surveys, particularly population-based surveys, are including a sexual orientation

[9]According to Boehmer et al. (2008), thoroughly constructed convenience samples will continue to constitute useful sampling strategies to further research on the LGBT population. This includes virtual methodologies to sample and recruit hard-to-reach LGBT populations (McDermott and Roen 2012), although evidence does suggest that differences exist between online and other data collection modalities (Reisner et al. 2014).

question that enables these subgroups to be separated out for analysis or comparison purposes. One example is the NHIS, a survey of the American population under the auspices of the U.S. Department of Health and Human Services (HHS). The NHIS is administered to about 40,000 households (approximately 100,000 individuals) annually in order to obtain a representative sample of the entire country. The NHIS started including questions about sexual orientation in 2013, but is postponing adding gender identity questions until they can convene a panel of experts to discuss the content of such questions (Brown 2010). Although few population-based surveys include questions on gender identity yet, this appears to be on the horizon for the NHIS (Brown 2010). Results from the 2013 NHIS regarding sexual orientation and health were published in July 2014 (Ward et al. 2014). Another large population-based survey that now includes sexual orientation and behavior questions and substance use questions is the National Health and Nutrition Examination Survey (NHANES). The NHANES is unique as it includes both interviews and individual examinations. Questions on smoking and tobacco use, alcohol use and drug use are included in the NHANES.

The Behavioral Risk Factor Surveillance System (BRFSS) is a statewide telephone survey examining behavioral risks among adults associated with premature morbidity and mortality. The BRFSS is the world's largest, ongoing telephone health survey system, is run by the Centers for Disease Control, and included cell phone surveys in the 2011 public release data set (Centers for Disease Control 2013). All fifty State Health Departments have been urged by the Fenway Institute's Center for Population Research in LGBT Health to include sexual and gender orientation questions on the BRFSS. The Fenway Institute reports 25 states and the District of Columbia included at least one item measuring a dimension of sexual orientation on the BRFSS between 1995 and 2011. Of these, 9 states included an item that assessed sexual orientation identity and an item that addressed same-sex behavior, 11 states and the District of Columbia included one item that assessed sexual orientation identity, and 5 states included one item that assess same-sex sexual behavior (although Georgia and Florida included these questions in only some counties) (Fenway Institute 2013a, b).[10]

Implications of sampling strategy are large, as poorly designed, small, nonrepresentative samples limit the generalizability of research. Despite progress, there continues to be a need for improved population-based sampling in LGBT research. Barriers exist, however, as many LGBT individuals are hesitant to "come out" and with good reason. Severe homophobic reactions have resulted in beatings and sometimes death for gay men and transgender women, in particular, particularly transgender women of color. While attitudes toward sexual and gender orientation have dramatically changed toward more acceptance, particularly among young people (Smith 2011), LGBT hate crimes are still common (45.3% of survivors and victims identified as gay, 20.6% as lesbian, 10.5% as transgender, and 8.7% as bisexual). In 2012 LGBT hate crime homicides, transgender women represented 50% of homicides and gay men 38.5%. Transgender women of color have been particularly hard hit (National Coalition of Anti-Violence Programs (NCAVP 2013). Alcohol and other drugs often play a role in these tragic events, although little research exists to describe their context.

[10]Additionally, the Fenway Institute's Center for Population Research in LGBT Health developed the Population Research in Sexual Minority Health (PRISM) Data Archive as a collaborative project with the Inter-university Consortium for Political and Social Research (ICPSR). This archive makes quality data sets available for analysis of LGBT issues in the U.S. Some of these data sets are only available to ICPSR members or member institutions. To download data, the institutional affiliation must be verified from a campus computer when setting up an account. After the account is verified, data may be accessed on a home computer. More information is available at the following site: http://www.icpsr.umich.edu/icpsrweb/FENWAY/datasets/.

16.4.4 Combining Sexual Orientation Groups for Analysis

While still relatively rare, there are an increasing number of large population-based studies of LGBT populations appearing in the research literature. Those that have been conducted indicate higher rates of ATOD use among certain, but not all sexual orientation groups within the LGBT spectrum. It is for this reason that the IOM recommended that these groups be analyzed separately and not combined into one group (IOM 2011). Combining of groups has been done often in the past due to small numbers of LGBT individuals self-identifying on questionnaires. However, more current research confirms the differences between the groups when analyzed separately. For example, combining lesbians and bisexual women into one group of sexual minority women (SMW) or lesbian/bisexual (LB) women for data analysis has led to the conclusion that lesbians and bisexual female college students have higher rates of smoking, binge drinking, and marijuana use as compared to their heterosexual counterparts. Studies that separate the groups, however, have sometimes found different results. In some studies, lesbian undergraduate college students were not significantly different than their heterosexual counterparts in these substance use behaviors (Eisenberg and Weschler 2003; Kerr et al. 2014; McCabe et al. 2005) while bisexual women were. In such cases, a combined group analysis of lesbians and bisexual women may result in the conclusion that lesbians have higher prevalence rates of these behaviors when in actuality they may not, while bisexual women actually do have higher rates (a Type 1 error).

Other categories may cause similar problems. In the end, simply defining groups only by their behaviors [e.g., women who have sex with women (WSW) or men who have sex with men (MSM)] combines gay and bisexual groups in a manner that may mask important differences between subgroups, with potentially negative repercussions for policy and practice solutions. These efforts also relegate sexual orientation to a behavior rather than a primary attraction or identity.

16.4.5 Ethical Considerations

The ethical considerations of researching substance abuse in LGBT populations are compounded. First there are the accepted ethical considerations associated with researching sensitive topics generally, as substance use and abuse is both socially sensitive and poses potential threat (e.g., legal sanction, etc.) to those involved (Lee 1993). However, substance abuse research conducted within the LGBT community is also being conducted on a population that is already socially stigmatized and marginalized as a sexual minority (Martin and Meezan 2003). This status can be even further complicated if LGBT subjects are also racial and ethnic minorities and/or minors, as the intersection of gender, age, race, and ethnicity further compounds marginalization (Wheeler 2003; Valentine et al. 2010). These statuses can often fundamentally shape both the conceptualization of research as well as the interactions between researcher and subject. Thus, researchers must take every precaution to ensure that subjects are not harmed.[11] This includes taking painstaking measures to ensure confidentiality and protection of human subjects, as LGBT research may require additional measures to ensure safety of study participants—both during and after the conduct of the study—and relevance of the study's findings (Martin and Meezan 2003). Recommendations include partnering with members of the target population in the design and conduct of the investigation, as well as exploration of innovative methodologies (Bettinger 2010; Martin and Meezan 2003; Wheeler 2003). In fact, considerations of study design are also paramount and range from understanding the terminology employed to defining the research population in light of the noted variation that exists within the LGBT community

[11]For a more detailed discussion, see Martin and Meezan's (2003) examination of the application of ethical standards (e.g., social work's Code of Ethics and psychology's Ethical Principles of Psychologists and Code of Conduct) to LGBT research. They also explore how bias may impact both the conceptualization and conduct of research.

(Bettinger 2010). The latter, in particular, challenges researchers to respect cultural differences, including language and identity preferences.

Sampling, recruitment and retention can also pose a number of ethical considerations. For instance, due to challenges associated with identification and location, LGBT studies may often employ a snowball method to sample LGBT members in a community. As this technique relies on interpersonal relations and connections between people, it both includes and excludes individuals, as well as informs on participants' private lives (Browne 2005; Penrod et al. 2003). Other purposive and convenience designs pose similar challenges, and researchers are encouraged to consider carefully the implications of sampling design before beginning a study.[12] Cost considerations abound in these considerations, but given the marginalized nature of the target population all efforts should be made to ensure their protection in the study process.[13]

16.5 Conclusion

Clearly, substance use and abuse are issues that must be addressed in the LGBT community. Research plays a critical role in understanding the problem and shaping evidence-based solutions. However, a major difficulty in conducting LGBT research is not only conducting the research in a culturally sensitive and appropriate manner, but in identifying individuals willing to participate. Most often, those willing are those who are already "out" such as members of LGBT organizations or pride groups, but these individuals may have very different attitudes and behaviors than those who are closeted and hiding their sexual or gender orientations. Thus, in order to determine more accurately what LGBT individuals believe or how they behave, a more inclusive and anonymous sampling procedure is needed, particularly for survey research. New and innovative research designs are also encouraged as a means to ensure adequate protection of subjects, while at the same time broadening the inquiry into the substance abuse problems impacting the community. Qualitative research methods may also add much needed detail about LGBT individual's lives.

Most importantly, more research is needed on substance abuse among all of the groups and subgroups in the LGBT spectrum. Additional funding streams are available to conduct this research now that LGBT health has been added to the Healthy People 2020 agenda. This research is particularly needed for bisexual and transgender individuals, not only due to documented health risk behaviors, but because they have been so long ignored. Researchers should utilize recommendations of the IOM Report and information from such resources as the Fenway and Williams Institutes when planning and conducting new research.

[12]While not discussed in detail here, qualitative study designs present further challenges, including sampling and addressing issues related to participant observation. For additional information, see Bettinger (2010) or Kelly (2010).

[13]The National Transgender Discrimination Survey is the most extensive survey of transgender discrimination ever completed. The survey team distributed their questionnaire through transgender-led or transgender-serving community-based organizations, organizations serving hard-to-reach populations and 150 active online community list serves. They also paid workers in homeless shelters, legal aid clinics, mobile health clinics, and other service settings to host "survey parties" using paper surveys. They conducted phone follow-ups for the paper surveys over three months. The majority of surveys were completed online and the final sample's geographic distribution mirrored that of the U.S. population. This survey research is an excellent example of developing and utilizing creative sampling techniques.

Appendix: Examples of National Data Sets Including Sexual Orientation and/or Gender Identity Questions

Data set name (abbreviation)/ Organization	Website access/contact person (notes)
American College Health Association-National College Health	None/Mary Hoban (mhoban@acha.org) (data must be requested

(continued)

Data set name (abbreviation)/ Organization	Website access/contact person (notes)
Assessment II (ACHA-NCHA-II)/ American College Health Association	on a form provided upon request—this is not a nationally representative sample)
Behavioral Risk Factor Surveillance System (BRFSS)/Centers for Disease Control and Prevention (CDC)	http://www.cdc.gov/brfss/data_documentation/ (sexual orientation questions included only in some states—currently 17)
National Health and Nutrition Examination Survey (NHANES)/CDC	http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm
National Health Interview Survey (NHIS)/CDC	http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm (sexual orientation question added in 2013; gender ID question not yet included)
National Survey on Drug Use and Health (NSDUH)/Substance Abuse and Mental Health Services Administration (SAMHSA)	http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/64 (sexual orientation and gender identity questions were added in 2015)
Population Research in Minority Health (PRISM)/Fenway Health (must be ICPSR member school to access this data set)	http://www.icpsr.umich.edu/icpsrweb/FENWAY/datasets/
Youth Risk Behavior Survey (YRBS)/CDC	http://www.cdc.gov/healthyyouth/data/yrbs/data.htm (sexual orientation questions recently included in 2015 YRBS representative sample)

References

Andrasik, M. P., Valentine, S. E., & Pantalone, D. W. (2013). Sometimes you just have to have a lot of bitter to make it sweet: Substance abuse and partner violence in the issues of HIV+ men who have sex with men. *Journal of Gay and Lesbian Social Services, 25,* 287–305.

AVERT. (2014). *HIV transmission routes among African Americans, sexual transmission of HIV*. East Sussex, UK: AVERT. Retrieved from: http://www.avert.org/hiv-aids-among-african%20americans.htm#footnote6_8gefl0x

Bettinger, T. V. (2010). Ethical and methodological complexities in research involving sexual minorities. *New Horizons in Adult Education and Human Resource Development, 24,* 43–58.

Boehmer, U., Clark, M., Timm, A., & Ozonoff, A. (2008). Tow means of sampling sexual minority women: How different are the samples of women? *Journal of LGBT Health Research, 4,* 143–151.

Boyd, C. J., McCabe, S. E., & d'Arcy, H. (2003). Ecstasy use among college undergraduates: Gender, race and sexual identity. *Journal of Substance Abuse Treatment, 24*(3), 209–215.

Brown, D. (2010, June 29). Main federal health survey will ask about sexual orientation, gender identity. *The Washington Post*. Retrieved from http://www.washingtonpost.com/national/health-science/main-federal-health-survey-will-ask-about-sexual-orientation-gender-identity/2011/06/29/AGMZwMrH_print.html

Browne, K. (2005). Snowball sampling: Using social networks to research non-heterosexual women. *International Journal of Social Research Methodology, 8,* 47–60.

Bux, D. A. (1996). The epidemiology of problem drinking in gay men and lesbians: A critical review. *Clinical Psychology Review, 16,* 277–298.

Centers for Disease Control. (2013, March 19). *About the behavioral risk factor surveillance system (BRFSS)*. Atlanta, GA: U.S. Centers for Disease Control and Prevention. Retrieved from http://www.cdc.gov/brfss/about/about_brfss.htm

Cochran, S. D., Ackerman, D., Mays, V. M., & Ross, M. W. (2004). Prevalence of nonmedical drug use and dependence among homosexually active men and women in the US population. *Addiction, 99,* 989–998.

Cochran, S. D., Keenan, C., Schober, C., & Mays, V. M. (2000). Estimates of alcohol use and clinical treatment needs among homosexually active men and women in the U.S. population. *Journal of Consulting and Clinical Psychology, 68,* 1062–1071.

Conron, K., Lombardi, E., & Reisner, S. (2014). Identifying transgender and other gender minority respondents on population-based surveys: Approaches. In J. L. Herman (Ed.), *Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys*. Los Angeles, CA: The Williams Institute.

Corliss, H. L., Rosario, M., Wypij, D., Wylie, S. A., Frazier, A. L., & Austin, S. B. (2009). Sexual orientation and drug use in a longitudinal cohort study of U.S. adolescents. *Addictive Behaviors, 35*(5), 517–521.

D'Augelli, A. R., Hershberger, S. L., & Pilkington, N. W. (2001). Suicidality patterns and sexual orientation-related factors among lesbian, gay, and bisexual youths. *Suicide and Life Threatening Behavior, 31,* 250–264.

Deutsch, M., Green, J., Keatley, J., Mayer, G., Hastings, J., & Hall, A. M. (2013). Electronic medical records and the transgender patient: Recommendations from the World Professional Association for Transgender Health EMR Working Group. *Journal of the American Medical Informatics Association, 20,* 700–703.

Diamant, A. L., Wold, C., Spritzer, K., & Gelberg, L. (2000). Health behaviors, health status, and access to and use of health care: A population-based study of lesbian, bisexual and heterosexual women. *Archives of Family Medicine, 9*(10), 1043–1051.

Drabble, L., Midanik, L. T., & Trocki, K. (2005). Reports of alcohol consumption and alcohol-related problems among homosexual, bisexual, and heterosexual respondents: Results from the 2000 National Alcohol Survey. *Journal of Studies on Alcohol, 66*(1), 111–120.

Drabble, L., & Trocki, K. (2005). Alcohol consumption, alcohol-related problems, and other substance use among lesbian and bisexual women. *Journal of Lesbian Studies, 9*(3), 19–30.

Eaton, L., Kaufman, M., Fuhrel, A., Cain, D., Cherry, C., Pope, H., et al. (2008). Examining factors co-exiting with interpersonal violence in lesbian relationships. *Journal of Family Violence, 23,* 697–705.

Eisenberg, M., & Weschler, H. (2003). Substance use behaviors among college students with same-sex and opposite-sex experience: Results from a national study. *Addictive Behaviors, 28*(5), 899–913.

Fifield, L. H., Latham, J. D., & Phillips, C. (1977). *Alcoholism in the gay community: The price of alienation, isolation, and oppression*. Los Angeles, CA: Gay Community Services Center.

Goldberg, S., Strutz, K. L., Herring, A. A., & Halpern, C. T. (2013). Risk of substance abuse and dependence among young adult sexual minority groups using a multidimensional measure of sexual orientation. *Public Health Reports, 128*(3), 144–152.

Grant, J. M., Mottet, L. A., Tanis, J., Herman, J., Harrison, J., & Keisling, M. (2010). *National Transgender Discrimination Survey Report on Health and Health Care*. Washington, DC: National Center for Transgender Equality/National Gay and Lesbian Task Force. Retrieved from http://transequality.org/PDFs/NTDSReportonHealth_final.pdf

Green, K. E., & Feinstein, B. A. (2012). Substance use in lesbian, gay, and bisexual populations: An update on empirical research and implications for treatment. *Psychology of Addictive Behaviors, 26,* 265–278.

Halkitis, P. N., & Palamar, J. J. (2008). Multivariate modeling of club drug use initiation among gay and bisexual men. *Substance Use & Misuse, 43,* 871–879.

Halkitis, P. N., Palamar, J. J., & Mukherjee, P. P. (2007). Poly-club-drug use among gay and bisexual men: A longitudinal analysis. *Drug and Alcohol Dependence, 89,* 153–160.

Hequembourg, A. L., Livingston, J. A., & Parks, K. A. (2013). Sexual victimization and associated risks among lesbian and bisexual women. *Violence Against Women, 19*(5), 634–657.

Hirshfield, S., Remien, R. H., & Chiasson, M. A. (2006). Crystal methamphetamine use among men who have sex with men: Results from two national online studies. *Journal of Gay & Lesbian Psychotherapy, 10,* 85–93.

Hughes, T. L. (2005). Alcohol use and alcohol-related problems among lesbians and gay men. *Annual Review of Nursing Research, 23,* 283–325.

Hughes, T. L., & Eliason, M. (2002). Substance use and abuse in lesbian, gay, bisexual and transgender populations. *The Journal of Primary Prevention, 22,* 263–298.

Hughes, T. L., McCabe, S. E., Wilsnack, S. C., West, B. T., & Boyd, C. J. (2010). Victimization and substance use disorders in a national sample of heterosexual and sexual minority women and men. *Addiction, 105*(12), 2130–2140.

Institute of Medicine. (2011). *The Health of lesbian, gay, bisexual, and transgender people: Building a foundation for better understanding*. Washington, DC: The National Academies Press.

Kelly, B. C. (2010). Sampling and recruitment issues in qualitative drugs research: Reflections on the study of club drug users in metro New York. *Substance Use and Misuse, 45,* 671–683.

Kelly, B. C., & Parsons, J. T. (2007). Prescription drug misuse among club drug-using young adults. *The American Journal of Drug and Alcohol Abuse, 33,* 875–884.

Kerr, D. L., Ding, K., & Chaya, J. (2014). Substance use of lesbian, gay, bisexual, and heterosexual college students. *American Journal of Health Behavior, 38*(6), 951–963.

Laumann, E. O., Gagnon, J., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago, IL: University of Chicago Press.

Lee, G. L., Griffin, G. K., & Melvin, C. L. (2009). Tobacco use among sexual minorities in the USA: 1987 to May 2007: A systematic review. *Tobacco Control, 18,* 275–282.

Lee, R. M. (1993). *Doing research on sensitive topics*. Thousand Oaks, CA: SAGE Publications.

Lewis, C. E., Saghir, M. T., & Robins, E. (1982). Drinking patterns in homosexual and heterosexual women. *Journal of Clinical Psychiatry, 43,* 277–279.

Lieb, S., Thompson, D. R., Misra, S., Gates, G. J., Duffus, W. A., Fallon, S. J., et al. (2009). Estimating populations of men who have sex with men in the Southern United States. *Journal of Urban Health, 86* (6), 887–901.

Lyons, T., Johnson, A. K., & Garofalo, R. (2013). "What could have been different": A qualitative study of syndemic theory and HIV prevention among young men who have sex with men. *Journal of HIV/AIDS & Social Services, 12,* 368–383.

Malebranche, D. (2011, April). The truth about the "Down Low". Washington, DC: American Psychological Association. Retrieved from http://www.apa.org/pi/aids/resources/exchange/2011/04/down-low.aspx

Martin, J. I., & Meezan, W. (2003). Applying ethical standards to research and evaluations involving lesbian, gay, bisexual, and transgender populations. In W. Meezan & J. I. Martin (Eds.), *Research methods with gay, lesbian, bisexual, and transgender populations*. New York, NY: Routledge.

McCabe, S. E., Bostwick, W. B., Hughes, T. L., West, B. T., & Boyd, C. J. (2010). The relationship between discrimination and substance use disorders among lesbian, gay, and bisexual adults in the United States. *American Journal of Public Health, 100*(10), 1946–1952.

McCabe, S. E., Boyd, C., Hughes, T. L., & dArcy, H. (2003). Sexual identity and substanceabuse among undergraduate students. *Substance Abuse, 24*(2), 77–91.

McCabe, S. E., Hughes, T. L., Bostwick, W., & Boyd, C. J. (2005). Assessment of difference in dimensions of sexual orientation: Implications for substance use research in a college-age population. *Journal of Studies on Alcohol and Drugs, 66,* 620–629.

McCabe, S. E., Hughes, T. L., Bostwick, W., Morales, M., & Boyd, C. J. (2012). Measurement of sexual identity in surveys: Implications for substance abuse research. *Archives of Sexual Behavior, 41,* 649–657.

McCabe, S. E., Hughes, T. L., Bostwick, W. B., West, B. T., & Boyd, C. J. (2009). Sexual orientation, substance use behaviors and substance dependence in the United States. *Addiction, 104*(8), 1333–1345.

McCrady, B. S. (2004). To have but one true friend: Implications for practice of research on alcohol use disorders and social networks. *Psychology of Addictive Behaviors, 18,* 113–121.

McDermott, E., & Roen, K. (2012). Youth on the virtual edge: Researching marginalized sexualities and genders online. *Quality Health Research, 22,* 560–570.

McKirnan, D. J., & Peterson, P. L. (1989). Alcohol and drug use among homosexual men and women: Epidemiology and population characteristics. *Addictive Behaviors, 14,* 545–553.

McNeil, J., Bailey, L., Ellis, S., Morton, J., & Regan, M. (2012). *Trans mental health study 2012*. Edinburgh, UK: The Scottish Transgender Alliance. Retrieved from http://www.gires.org.uk/assets/Medpro-Assets/trans_mh_study.pdf

Measham, F., Wood, D. M., Dargan, P. I., & Moore, K. (2011). The rise in legal highs: Prevalence and patterns in the use of illegal drugs and first- and second-generation 'legal highs' in South London gay dance clubs. *Journal of Substance Use, 16,* 263–272.

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129,* 674–697.

Molina, Y., & Ramirez-Valles, J. (2013). HIV/AIDS stigma: Measurement and relationships to psycho-behavioral factors in Latino gay/bisexual men and transgender women. *AIDS Care: Psychological and Socio-Medical Aspects of AIDS/HIV, 25*(12), 1559–1568.

Mosher, W. D., Chandra, A., & Jones, J. (2005). Sexual behavior and selected health measures: Men and women 15–44 years of age, United States, 2002. *Advance Data, 362,* 1–55.

National Coalition of Anti-Violence Prevention Programs (NCAVP). (2013). *Lesbian, gay bisexual, transgender, queer and HIV-affected hate violence in 2012*. New York, NY: National Coalition of Anti-Violence Programs. Retrieved from http://www.avp.org/storage/documents/ncavp_2012_hvreport_final.pdf

Newcomb, M., Birkett, M., Corliss, H., & Mustanski, B. (2014). Sexual orientation, gender, racial differences in illicit drug use in a sample of US high school students. *American Journal of Public Health, 104,* 304–310.

Parsons, J. T., Kelly, B. C., & Wells, B. E. (2006). Differences in club drug use between heterosexual and lesbian/bisexual females. *Addictive Behaviors, 31*(13), 2344–2349.

Penrod, J., Preston, D. B., Cain, R. E., & Starks, M. T. (2003). A discussion of chain referral as a method of sampling hard-to-reach populations. *Journal of Transcultural Nursing, 14,* 100–107.

Purcell, D. W., Moss, S., Remien, R. H., Woods, W. J., & Parsons, J. T. (2005). Illicit substance use, sexual risk, and HIV-positive gay and bisexual men: Differences by serostatus of casual partners. *AIDS, 19,* S37–S47.

Rhodes, S. D., Hergenrather, K. C., Yee, L. J., Knipper, E., Wilkin, A. M., & Omli, M. R. (2007). Characteristics of a sample of men who have sex with men, recruited from gay bars and internet chat rooms, who report methamphetamine use. *AIDS Patient Care and STDS, 21,* 575–583.

Reisner, S. L., Conron, K., Scout, N., Mimiaga, M. J., Haneuse, S., & Austin, S. B. (2014). Comparing in-person and online survey respondents in the U.S. National Transgender Discrimination Survey: Implications for transgender health research. *LGBT Health, 1,* 98–106.

Rukus, J., Stogner, J., & Miller, B. (2016). LGBT novel drug use as contextualized through control, strain, and learning theories. *Social Science Quarterly*. doi:10.1111/ssqu.12329.

Ryan, C., Russell, S. T., Huebner, D., Diaz, R., & Sanchez, J. (2010). Family acceptance in adolescence and the health of LGBT young adults. *Journal of Child and Adolescent Psychiatric Nursing, 23*(4), 205–213.

San Francisco Department of Public Health. (1999). The Transgender Community Health Project, HIV InSite. San Francisco, CA: University of California. Retrieved from http://hivinsite.ucsf.edu/InSite?page=cftg-02-02#S4.7X

Savin-Williams, R. C. (2006). Who's gay? Does it matter? *Current Directions in Psychological Science, 15*(1), 40–44.

Scheer, S., Peterson, I., Page-Shafer, K., Delgado, V., Gleghorn, A., Ruiz, J., et al. (2002). Sexual and drug use behavior among women who have sex with both women

and men: results of a population-based survey. *American Journal of Public Health, 92*(7), 1110–1112.

Smith, T. W. (2011). *Public attitudes toward homosexuality*. Chicago, IL: NORC/University of Chicago. Retrieved from http://www.norc.org/NewsEventsPublications/PressReleases/Pages/american-acceptance-of-homosexuality-gss-report.aspx

Stall, R., Paul, J. P., Greenwood, G., Pollack, L. M., Bein, E., Crosby, G. M., et al. (2001). Alcohol use, drug use and alcohol-related problems among men who have sex with men: The Urban Men's Health Study. *Addiction, 96,* 1589–1601.

The Fenway Institute. (2013a, July 29). *25 states have included sexual orientation items in the BRFSS*. Boston, MA: The Fenway Institute. Retrieved from http://lgbtpopulationcenter.org/2013/07/25-states-have-included-sexual-orientation

The Fenway Institute. (2013b, December). *Asking patients questions about sexual orientation and gender identity in clinical settings: A study in four health centers*. Boston, MA: The Fenway Institute. Retrieved from http://thefenwayinstitute.org/wp-content/uploads/COM228_SOGI_CHARN_WhitePaper.pdf

The GenIUSS Group. (2014). *Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys*. J. L. Herman (Ed.). Los Angeles, CA: The Williams Institute. Retrieved from http://williamsinstitute.law.ucla.edu/wp-content/uploads/geniuss-report-sep-2014.pdf

Thompson, E. M., & Morgan, E. M. (2008). "Mostly straight" young women: Variations in sexual behavior and identity development. *Developmental Psychology, 44,* 15–21.

Trocki, K. F., Drabble, L. A., & Midanik, L. T. (2009). Tobacco, marijuana, and sensation seeking: Comparisons across gay, lesbian, bisexual, and heterosexual groups. *Psychology of Addictive Behaviors, 23*(4), 620–631.

Valanis, B. G., Bowen, D. J., Bassford, T., Whitlock, E., Charney, P., & Carter, R. A. (2000). Sexual orientation and health: Comparisons in the women's health initiative sample. *Archives of Family Medicine, 9*, 843–853.

Valentine, G., Butler, R., & Skelton, T. (2010). The ethical and methodological complexities of doing research with 'vulnerable' young people. *Ethics, Place & Environment, 4,* 119–125.

Volk, J. E., Prestage, G., Jin, F., Kaldor, J., Ellard, J., Kippax, S., & Grulich, A. E. (2006). Risk factors for HIV seroconversion in homosexual men in Australia. *Sexual Health, 3*(1), 45–51.

Ward, B. W., Dahlhamer, J. M., Galinsky, A. M., & Joestl, S. S. (2014). Sexual orientation and health among U.S. adults: National Health Interview Survey, 2013. *National Health Statistics Reports, 77,* 1–10.

Wheeler, D. P. (2003). Methodological issues in conducting community-based health and social services research among urban Black and African-American LGBT populations. *Journal of Gay & Lesbian Social Services, 15,* 65–77.

Ziebold, T. O., & Mongeon, J. E. (1982). Introduction: Alcoholism and the homosexual community. *Journal of Homosexuality, 7,* 3–7.

Conducting Substance Abuse Research: Incarcerated Populations

17

Sage Kim and Michael Puisis

17.1 Introduction

Conducting substance abuse research often leads to studying correctional populations, as the majority of the 2.2 million persons in both prisons and jails had regularly used drugs prior to their incarceration, and drug use is one of the factors for current mass incarceration (Carson 2015; Minton and Zeng 2015; Mumola and Karberg 2006). Drug policies implemented since the 1970s have contributed to the increasing number of people in jails and prisons, including mandatory minimum sentences, three strikes laws, and truth-in-sentencing requirements (Bobo and Thompson 2010; Mackenzie 2001; Mumola and Beck 1997). Consequently, researchers interested in exploring the patterns of substance abuse and related social, behavioral, and health problems often turn their attention to incarcerated populations. Correctional settings, however, introduce unique challenges for researchers.

Security is the primary concern of jails and prisons, as such, these facilities are tightly controlled and continuously monitored. Typically, inmates' movement is restricted, and men and women live in separate housing units. Inmates are also separated based on classification schemes that include gang affiliation, mental health conditions, medical conditions, and propensity for violence. For researchers, electronic devices such as computers, phones, and cameras are, for the most part, prohibited in correctional facilities. The primacy of security affects all aspects of research in the corrections. In addition to formal institutional review board (IRB) approvals, researchers in correctional facilities typically need to gain approval from correctional authorities and/or the health authority within a correctional facility. Unless proposed research directly benefit existing institution's programs or needs, correctional authorities may be reluctant to permit access to the facility. Furthermore, logistics of gaining access to inmate populations is also difficult. Inmates have breakfast as early as 3:30 in the morning, are locked in their cells to be counted several times a day, and may have scheduled court dates, which limit available free time. Interviews can easily be canceled when other correctional activities take place, or if an officer simply forgets to bring designated inmates. There are also emergency situations that result in facility-wide lockdown, in which case researchers may not be able to carry out scheduled interviews, or may not be able to get in or out of the facility. These types of situations can be frustrating, but figuring out effective ways to work around the highly structured (and simultaneously unpredictable) correctional activities is necessary (Blagden and Pemberton 2010).

Inmates may share norms, culture, and/or personal characteristics, many of which can be specific to jail/prison settings. These potential

S. Kim (✉)
School of Public Health, University of Illinois at Chicago, Chicago, IL, USA
e-mail: skim49@uic.edu

M. Puisis
Correctional Consultant, Chicago, IL, USA
e-mail: mpuisis@gmail.com

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_17

differences may affect study participants' responses to research questions, and therefore need to be examined while recruiting, collecting data, and implementing interventions. Furthermore, inmates are confined against their will, and many mistrust authority and the system, feeling threatened by the correctional system, as well as the officers. Oftentimes researchers can be perceived as part of this system, resulting in mistrust and suspicion. Thus, it is important to communicate clearly to all participants not just research objectives, but also confidentiality and privacy considerations.

More often than not, researchers are not employees of the correctional system, and as outsiders, can be seen as a potential security risk. Researchers often spend an enormous amount of time and energy gaining the trust and cooperation of correctional officers and other administrative personnel. Correctional officials may be reluctant to allow outside researchers to come in and 'study' the correctional system, which has the potential to uncover misconduct or other issues. For the correctional authority, it is an unnecessary burden to risk potentially negative study findings that could cause problems for them down the road (Fox et al. 2011).

Researchers working in correctional settings need to interact with a variety of stakeholders (Fox et al. 2011; Freudenberg 2007). Even after correctional authority and IRB approve a research project, ground-level staff and officers do not necessarily need to accommodate researchers. Many already feel that they are overburdened, working in an ever-expanding and overcrowded facility. As such, they may feel that a research project would only add to their already heavy workload. Worse yet, they may discourage inmates from participating (Freudenberg 2007). Others might be reluctant to facilitate research activities in their unit or service areas (Blagden and Pemberton 2010). Overall, however, correctional staff members are willing to cooperate with researchers when they understand and support research objectives. Establishing collaboration with healthcare providers, officers, educators, and other entities can help to not only overcome initial resistance, but ultimately tap into insiders' intimate knowledge of nuanced situations and meaning (Megargee 1995).

Oftentimes, it is difficult to evaluate the success of interventions in correctional facilities, partly due to the fact that substance abuse research requires a long-term view of outcome measures, which often cannot be captured while study participants are still confined in jail/prison (Freudenberg 2007). Furthermore, key elements of successful substance abuse intervention research cannot be directly dealt with within correctional settings, such as social support, employment, stable housing, adjustment of social networks, and access to health and social services. Studies often require follow-up after inmates return to the community and measure how former inmates deal with factors that affect their drug use triggers, behavior, and decision-making (Dennis et al. 2000). Studies attempting to follow-up inmates after release frequently encounter considerable difficulty, most notably with high attrition rates. Many inmates come from disadvantaged neighborhoods or live in unstable housing situations. Frequent drug relapse, recidivism, and the illegal nature of drug and crime presents additional challenges to reconnecting with former inmates. Moreover, the conditions that led to drug use in the community in the first place are not easily manipulated, and often former inmates return to previous risky behavior after release from custody. Thus determining measurable outcomes and time frame is important for research in correctional setting.

There are many challenges for research conducted in correctional facilities, including threats to reliability, as well as internal and external validity (Cook and Campbell 1979; Dennis et al. 2000; Fletcher and Tims 1992). Because of the unique conditions within which correctional research takes place, potential pitfalls associated with design, methods, implementation, and interpretation of research with incarcerated populations can be much more pronounced and difficult to overcome (Fletcher and Tims 1992; Pelissier et al. 2007). The following chapter identifies some of the challenges related to study design, measurement, and evaluation of substance abuse research in correctional settings.

Strategies to overcome and advance research are presented, addressing the physical, psychosocial, and behavioral health of inmates and the community.

17.2 Data Collection

Study participants may intentionally and/or unintentionally under-report their substance use and other related health and behavioral measures. Imperfect memory, telescoping, social desirability, or legal implications can affect participants' responses (Junger-Tas and Haen Marshall 1999). Underreporting of substance use can be problematic as inmates are often incarcerated for drug-related charges, and might fear that their study responses could affect subsequent court proceedings and sentencing outcomes.

The issue of over-reporting can equally affect responses, as telescoping, or remembering only select events, can cause over-reporting (Fletcher and Tims 1992). Furthermore, individuals who seek a sense of power or 'respect' may exaggerate their engagement with drugs or criminal activities; equally misleading, they may fashion their answers to gain sympathy of researchers, in hopes of influencing their court hearing, to access medications (Lennox and Dennis 1994), or even for better food choices (Peters et al. 2005; Sierles 1984; Wilson 2000).

Despite conflicting results concerning self-reported drug use and delinquent behavior (Harrison 1997; Harrison et al. 2007; Richter and Johnson 2001), studies have documented that self-reported data can be reliable (Cook et al. 1997; Sutton et al. 2011); particularly, self-reported data concerning crime and victimization, which have been shown to be quite accurate (Webb et al. 2006).

Researchers will have a difficult time verifying self-reported data. While laboratory tests can confirm biomedical measures, many social science measures cannot be verified (Fletcher and Tims 1992). Furthermore, lab results can introduce false positive or false negative errors, which may lead to a false sense of security for researchers (Dennis et al. 2000). More importantly, inmates are unlikely to agree to drug testing in an environment where such an activity is illegal, subjecting them to harsher sentencing or punishment in the event of a positive result.

17.3 Data Quality and Response

Previous studies have illustrated that social desirability can affect participants' responses (Johnson et al. 2005a, b; Krumpal 2013). This, however, may be more prevalent when conducting research that explores deviance and/or crime. Inmates may seek to provide responses that are more socially acceptable, which could lead to an under-reporting of substance use and their role in criminal activities, or lead to over-reporting of their position in gangs, incarceration history, or number of sex partners.

Inmates often perceive researchers as part of the correctional system, and frequently ask for favors concerning their housing, food, or medications, even after an extensive explanation as to how the researchers are not part of the correctional facility's operation. This posits a potential bias, as their survey responses are influenced by how they perceive researchers' position in the correctional system, and they might anticipate what researchers are looking for (Bosworth et al. 2005; Fletcher and Tims 1992; Gostin et al. 2006). Many times, inmates provide responses expecting to gain benefits. For example, if an inmate believes that giving a certain response will result in housing considerations within the correctional facility, they may give that response. This requires careful questionnaire design and introductory explanations of study protocols. Researchers may need to ask same questions in different ways in order to arrive at the individual's true response. Because of the inevitable power dynamics between inmates, officers, and researchers, potential for this type of bias might be much more frequent in studies in correctional settings compared with studies in the general population, thereby presenting a threat to internal validity (Blagden and Pemberton 2010; Liebling 2001). Communicating research protocols and goals clearly can help to overcome such bias. But

the desire to establish closer relationships with inmates as a means to obtain more accurate information sometimes puts researchers at risk. Additionally, researchers based in correctional settings often encounter requests for help from former inmates in dire situations. Many times, these former inmates contact research personnel for help even after the research has ended (Blagden and Pemberton 2010). It is difficult to argue for or against researcher's engagement beyond the scope of research, especially when researchers are fully aware of difficult life conditions of study participants. Furthermore, while researchers are expected to be value-neutral, the aforementioned factors could affect how researchers view repeat offenders. Despite the desire to assist individuals in difficult situations, researchers must maintain neutrality and focus on their specific research questions.

Some researchers are criticized as being "drive-by" ethnographers, which refers to the collection of necessary data, followed by a swift exit. This type of research approach lacks long-term commitment and in-depth understanding of both people and place, and breeds mistrust; while potentially true of any research, it is more poignant in the incarcerated population (Blagden and Pemberton 2010; Bundy 2004; Cowburn 2010).

It is quite difficult to reconnect with former jail inmates after their release, and often follow-up interviews and interventions suffer from high attrition rates (Peters et al. 2005). Attrition can be a serious issue with studies attempting to conduct multi-time point, repeated measures, both in jail and after release. Individuals more likely to drop out of research are those who relapse, are re-incarcerated, or are living in unstable conditions due to various other reasons. Thus, attrition in studies with incarcerated individuals may systematically affect study outcomes (Abrams 2010; Fletcher and Tims 1992; Gelberg et al. 2000).

17.4 Voluntary Participation

Selection bias is a problem for any research project, but may be even more problematic for correctional research. Previous studies have argued that offenders who voluntarily participate in research have substantially different characteristics than those who choose not to participate (Harkins and Beech 2007; Marques et al. 2005; Megargee 1995). Some inmates might be less comfortable participating in studies because they worry if their responses negatively affect their stay in jail, in which case researchers should be aware that study findings could be biased toward an underestimation of a phenomenon being studied. In addition, sampling bias across different dorms/housing arrangements within a correctional facility could also affect study findings. Typically, housing units for lower security inmates are more accessible to a researcher than a high-security housing unit. Since inmates are separated based on sex, severity of alleged offenses or conviction, medical illness or psychiatric illness, researchers could inadvertently exclude some inmates who are different from other inmates. This potential for systemic selection bias in sampling should be carefully considered.

On the other hand, some inmates are more likely to be included in studies, because they are already ordered to medical care or other treatment programs by clinicians or drug judges. Substance use researchers in correctional settings often encounter inmates through these treatment units or programs. When evaluating an intervention in this scenario, researchers need to be careful not to overestimate treatment effects, as there is potential for regression toward the mean.

Participation in substance abuse studies would do well by not focusing solely on incarcerated populations, accounting for potential differences between drug use in incarcerated populations, as opposed to drug use in civilian non-incarcerated populations. Moreover, researchers need to be aware that different racial/ethnic groups are affected differently by present drug laws (Megargee 1995), and the population of substance

abusers in correctional settings may not be reflective of the overall drug using population in the United States. For example, minorities, particularly African Americans, are overrepresented in correctional settings (The Sentencing Project 2014). Thus, the distributions of substance abuse and related co-occurring conditions in jails and prisons cannot be directly generalized to the general population. Carefully describing how substance users in corrections fit with the overall substance abuse problem can help to address this problem.

17.5 IRB, Privacy, and Confidentiality

Inmates in correctional facilities are considered a special class with respect to IRB regulations. For inmates in particular, ensuring voluntary participation, privacy, and confidentiality is vital. Although it is unlikely that researchers would intentionally attempt to coerce inmates into research participation, the issue of coercion still needs to be carefully examined. As inmates are held against their will, without freedom of movement, they are particularly vulnerable to coercion (Blagden and Pemberton 2010; Freudenberg 2007; Gostin et al. 2006; Megargee 1995).

Privacy can be easily compromised in correctional settings, and arranging private interviews can be difficult. Inmates who agree to participate need to be moved by officers, called out from common rooms or dorms where everyone else can observe what they are doing. An inadvertent breach of confidentiality can also be an issue. For example, many healthcare providers report that dispensing certain types of medications can indicate what type of illness those who take the medications have (Wakeman and Rich 2010). Inmates often talk to other inmates about their engagement with research, and details of studies quickly become shared knowledge. Confidentiality of personal information must be carefully guarded with respect to other inmates.

IRB have a special provision for 'prisoner research' that requires a prisoner representative to review research protocols (Gostin et al. 2006). IRB review for research protocols involving inmates can be a lengthy process, and often require multiple revisions and clarifications (Freudenberg 2007). IRBs sometimes have difficulty recruiting prisoner representatives, in which case reviews are delayed. To complicate the process even further, many IRB reviewers are not familiar with research involving inmates; long, complicated, and frequently uncertain situations that are inevitable in research with inmates can be difficult to endure. Obtaining a Certificate of Confidentiality from the federal government can provide an additional protection.

Researchers often encounter situations in which protection of confidentiality, security, and safety of participants, as well as others, is difficult. Moreover, ethical concerns while conducting research with inmates seem to be much more complex than IRB protocols can address.

17.6 Intervention Effects in Jail and Generalizability

With respect to internal and external validity, research conducted in correctional settings presents similar difficulties compared to non-correctional studies, although the characteristics of inmates and correctional facilities often intensify these methodological challenges (Fletcher and Tims 1992). Correctional environments can cause tension between inmates and officers (Homel and Thomson 2005; Muscat 2008). Furthermore, inmates do not have personal space to avoid potential conflicts. The amount of drug use within correctional facilities is unknown. The expectation is that obtaining illegal substances is more difficult while incarcerated than when in the community, and this may create physical and/or psychological tension. Oftentimes, inmates experience physical ailments related to sudden withdrawal from drugs and withdrawal symptoms need to be effectively dealt with upon intake (Binswanger et al. 2009; Federal Bureau of Prisons Clinical Practice Guidelines (FBOP) 2014; Maruschak 2006).

A Bureau of Justice Statistics survey in 2004 reported that "53% of State and 45% of Federal prisoners met the DSM-IV criteria for drug

dependence or abuse" (Mumola and Karberg 2006). Similarly, 68% of jail inmates were dependent or abusing drugs or alcohol in 2002 (Karberg and James 2005). Because of the high prevalence rates of substance use/abuse among incarcerated individuals (Fazel and Danesh 2002; Mumola and Karberg 2006; Sedlak and McPherson 2010; The GAINS Center 2004), substance abuse screening is a necessary part of intake and clinical evaluation in correctional settings.

Correctional settings can provide unique opportunities for screening and initiating drug treatment programs, and a number of studies have documented the effectiveness of substance use treatment among incarcerated individuals (Du et al. 2013; Evans et al. 2006; Lipton 1994; Peters et al. 2000). However, substance abuse treatment recommendations are not specific to incarcerated populations, and thus the effectiveness of correctional treatment programs can be hard to estimate. Within correctional facilities, inmates are artificially isolated from their usual environment and events that trigger drug use. Highly structured living situations in correctional facilities and supportive interactions from treatment providers and peers contribute to successful treatment outcomes, but these support systems may not exist when they are released (La Vigne et al. 2008; Osher et al. 2003).

Despite its complex intersecting causes, interventions for inmates with substance use problems focus predominantly on individual level behavioral change (Ford and Airhihenbuwa 2010; Thomas et al. 2011). These individual level interventions provide self-monitoring skills to identify and modify criminal thinking patterns and substance abuse behavior, such as denial, externalization, and other distorted thought patterns (Carpenter 2001; McHugh et al. 2010; Peters et al. 2005). However, without taking into consideration contextual readiness, neighborhood level resources, and access to care, individual level readiness measures only offer a limited assessment of one's ability to change (Butzin et al. 2002; Grella and Greenwell 2007; Matheson et al. 2011). Selecting appropriate outcome measures for treatment programs implemented in corrections is a difficult task, in part because of ambiguous outcome measures and endpoints of interventions. Increasingly, researchers are interested in exploring effects of longer term interventions, and many correctional-based drug treatment programs include elements of assisting inmates' transition to community settings, which is a critical time for successful reentry. Still, hands off between movements and transition can lead to high dropout rates, an increased risk of resuming drug use, and higher probability of overdose immediately upon release (Binswanger et al. 2007, 2012). Despite discharge planning and transitional care, fragmented services between correctional settings and community organizations make it difficult to coordinate and ensure continuity of care for individuals leaving jails and prisons (Community Oriented Correctional Health Services (COCHS) 2010; The Centers for Disease Control and Prevention (CDC) 2001).

Researchers must set a proper time frame to assess intervention effects (Lennox and Dennis 1994). Substance abuse is now accepted as a chronic disease, which requires long-term treatment and maintenance. But because of the relatively short length of stay in jail, it is difficult to assess long-term treatment effects. This can be overcome by combining research within the correctional setting with community follow-up. Although it is well documented that reconnecting with inmates after they return to the community is difficult, in part because many inmates go back to their unstable housing conditions or relapse to drug abuse (Blandford and Osher 2013; Substance Abuse and Mental Health Services Administration (SAMHSA) 2005).

17.7 Measurement Issues

Research in correctional settings may require additional validation of measurement tools that have been originally developed for the general population (Dennis et al. 2000; Johnson et al. 2006; Megargee 1995), as the characteristics of correctional settings may influence how inmates perceive and assess their psychosocial conditions

when asked, which may alter inmates' responses (Steiner 2008). Additionally, inmates' life circumstances prior to incarceration may need to be closely examined when evaluating measurement instruments. Scholars have argued that a myriad of factors may affect how subgroups of individuals respond to health outcomes assessment tools, particularly because of the ways they may interpret questions and the meaning of the words used (Hahn and Cella 2003; Johnson et al. 2005a, b; Kagawa-Singer 2000; Stewart and Nápoles--Springer 2003). Measurements that are often developed for a dominant group, or the general population, inherently reflect a dominant culture and perspectives, and may not reflect life experiences of inmates who are often excluded from mainstream society (Megargee 1995). Scholars have argued that cross-cultural validation of measurements is required in order to gain accurate data (Hahn and Cella 2003; Stewart and Nápoles-Springer 2000). The quality of measurements must then be evaluated within incarcerated populations, and should be revised to accurately capture the psychosocial and behavioral experiences of inmates.

The length of survey questions also needs to be adjusted for correctional environment, as limited access to inmates and controlled environment makes it difficult to complete large questionnaires, and will not be particularly appreciated by corrections staff. While researchers should not cut down questions just to accommodate correctional staff, they must plan in advance both how and when they will conduct interviews, in order to integrate effectively into daily operations of correctional facilities. In the general population, parsimonious instruments are valued, but in the corrections field, it might be even more important to be able to collect necessary data without interfering with correctional schedules.

Considering literacy level is also an important element in designing effective instruments. Many inmates have low reading levels (Greenberg et al. 2007) and their ability to focus on one task may be affected by substance use and other co-occurring conditions, which need to be taken into account when designing and implementing questionnaires. Oftentimes, researchers are interested in participants' history of substance use and incarceration, and studies suggest that inmate interviews are as reliable as non-correctional settings (Newman 1958). However, questions about criminal charges may lead to inaccurate answers when inmates feel that their responses may affect sentencing results (Newman 1958; Schlosser 2008). Also, it is difficult to use medical records or correctional databases to confirm self-reported information, in part because the use of medical records must conform to HIPAA guidelines; linking medical and correctional data requires additional approvals from custody authorities.

One way to improve data quality is to utilize a mixed methods approach. When study participants are given a chance to talk about a series of related events, they often recall information more accurately (Weiss 1994). Interview probes and associations of events described by participants can provide ways to detect discrepancies in participant's statements, allowing researchers to gently ask for clarification (Reimer and Matthes 2007; Weiss 1994). Combining surveys and qualitative interviews can improve the credibility of self-reported data (Thomas et al. 2011). Using mixed methods is not only useful to improve validity of self-reported data, but also to examine complex intersecting factors contributing to both drug abuse and incarceration (Dennis et al. 1994).

17.8 Non-recursive, Multilevel, and Nonlinear Models

Health disparities studies have documented that broader social, economic, and political factors profoundly shape one's physical, mental, and behavioral health. Individuals living in disadvantaged neighborhoods are disproportionately exposed to social, economic, and environmental risks, and consequently, are more likely to be affected by physical, mental, and environmental health problems (Smedley and Myers 2014). This structural inequality has individual, interpersonal, and macro-level consequences. Issues

of racialized incarceration and drug policies cannot be ignored when developing effective interventions for the incarcerated (Crenshaw 1991; Sampson et al. 1997; Sampson and Jeghim-Bartusch 1998; Sharkey and Sampson 2010; Thomas and Sampson 2005; Wacquant 2009, 2010). Multiple co-occurring disorders and disadvantages should be considered in the analysis. For example, HIV and STI rates are higher among inmates compared with the general population (Centers for Disease Control and Prevention 2012). Similarly, mental health problems and homelessness are much more frequent among inmates than the general population (Greenberg and Rosenbheck 2008).

The relationships between individual socioeconomic conditions, neighborhood disadvantage, health, substance use, and incarceration are non-recursive in nature. However, more traditional theoretical models applied to substance abuse and incarceration research, such as social disorganization theory or the basic systemic model of crime, only allow unidirectional causal relations, and are thus inadequate to estimate reciprocal causal effects of neighborhood context, substance abuse, and incarceration.

In recent years, many health disparities research findings document that context matters, and individual as well as neighborhood socioeconomic disadvantage contributes to one's stress level, mental health, substance use, and incarceration (Draine et al. 2002; Greenberg and Rosenbheck 2008; Ross 2000; Ross and Mirowsky 2001). Most inmates return to the community, and many inmates are caught in situations where they constantly cycle through jail and poverty-stricken communities (Clear et al. 2003; Rose and Clear 1998; Sampson 2010; Sampson and Groves 1989). Poverty and joblessness rates are also much higher among inmates than in the general population (Mauer and King 2007). Substance abuse issues in correctional facilities are not just a cognitive, behavioral, or moral concern, but perhaps a consequence of uneven distribution of individual and neighborhood level risk exposure and disadvantage that can only be fully understood when approached with broader equality and justice perspectives (Brewer and Heitzeg 2008; Renauer et al. 2006; Rocque 2011). Researchers have argued that neighborhood institutions expand one's social support and access to resources by linking and bridging between individuals and organizations (Rose and Clear 2002; Sampson et al. 2000; Small 2009). Thus, multilevel modeling may help explain individual factors nested within contextual factors (Duncan et al. 2002; Kreft 1994; Miller and Moulton 2014).

Research concerning substance use and incarceration/crime may require statistical methods that combine statistical approaches, such as multilevel, geographic, and structural equation models, to establish a broader understanding of substance abuse and incarceration. From a measurement standpoint, multiple related explanatory variables of substance abuse and co-occurring conditions confound the underlying relationships between predictors and outcomes (Adlera et al. 2012; Evans and Kim 2010). Individuals are embedded within particular physical and social contexts, which affect norms, beliefs, and behavior (Krieger and Smith 2004; Subramaniana et al. 2001; Susser and Susser 1996). Scholars have used multilevel hierarchical models to understand individual and neighborhood level factors affecting substance use and crime (Boardman et al. 2001; Duncan et al. 2003; Stockdalea et al. 2007). Disadvantaged neighborhoods are more likely to experience physical and social disorganization, which leads to crime and incarceration (Bursik and Grasmick 1993; Shaw and McKay 1942), and more crime and incarceration damage stability of these neighborhoods, eroding collective capacity to deal with neighborhood problems. Social capital and collective efficacy are concepts that describe access to resources and ability to curb disorder at the community level (Browning et al. 2004; Small 2008). Research attempting to understand mechanisms of substance use and its relation to crime and incarceration cannot afford to ignore contextual factors. Research in correctional settings, in fact, is a critical part of understanding

society as a whole, particularly as to how a society defines what crime is, and who needs to be excluded from mainstream society (Thomas et al. 2011). Corrections serve as a tool to exclude, physically separating a select group of people from the public (Foster and Hagan 2007; Wacquant 2009, 2010; Wakefield and Uggen 2010).

The exponential increase in the incarceration rate since the early 1990s, predominantly among racial/ethnic minorities, is closely associated with racialized law enforcement practices, such as the Anti-Drug Abuse Acts of 1986 (Coyle 2002; The Sentencing Project 2013; Wacquant 2010). Furthermore, urban poverty areas are disproportionately affected by excess incarceration. For example, studies have documented that over 32% of inmates in Chicago come from only 7 of 77 community areas in Chicago, and these areas are characterized by high racial residential segregation, poverty, and lack of economic potential (Alfred and Chlup 2009; Morenoff and Sampson 1997; Sampson 2010; Sampson et al. 1997; Western 2006; Western et al. 2001). The use of spatial analysis and geographic information systems can help to pinpoint areas that are significantly affected by current incarceration practices (Dankwa-Mullan et al. 2010; Thomas et al. 2011).

Many issues that inmates face in relation to substance use further compound the effects of cumulative disadvantage over time. For example, studies have shown that a large proportion of female inmates had a history of childhood sexual abuse, which seems to affect depression and other mental health issues, school dropout, and eventually substance use (Messina and Grella 2006; Paone et al. 2009). Female inmates' pathways to drugs and crime also often entail involvement with older men who provided for them, but concurrently were abusive. Multiple related adverse events in inmates' lives may be better explored with structural equation modeling, which helps to unveil pathways and mediating factors between childhood traumatic events and substance use and incarceration (Bollen 2002; Furr and Bacharach 2008).

17.9 Politics, Funding Support, and Continuity of Research

Crime, and in many ways disease, is a socially constructed concept (Goode and Ben-Yehuda 1994; Potter 1996; Victor 1998). Crime delineates what is acceptable from deviant within a set of social norms (Becker 1997; Kitsuse and Cicourel 1963), and, as any other socially constructed categories, the definitions of deviance, as well as substance abuse, are prone to change over time. For example, the use of marijuana, once illegal across the country, is now becoming legal in several states.

And along with the ideas of crime and punishment, scholarly theories and best practices have also evolved over time. Whether rehabilitative or punitive, public discourse concerning public safety and fear of crime affects ways in which crime and offenders are managed (King and Mauer 2002; Mauer 2011). "Tough on crime" and "war on drug" policies are examples of this. While the public might feel that while offenders are behind bars they will not commit crime, which may imply better public safety, the long-term negative effects of excess incarceration have been well documented (Mauer 2011; Stevenson 2011; Thomas and Torrone 2006).

In a historical review of criminological perspectives over the past 40 years, Garland suggests that a complex set of institutional changes has occurred, which gave rise to a new form of crime control, including the decline of the penal welfare and rehabilitation programs, and the reemergence of punitive discourse and fear of crime as a contemporary cultural theme. Language of crime control is also an important indicator, which is a reflection of a set of practices of the criminal justice system, and largely dependent on external social structure and cultural norms. Garland argues that along with the emergence of control theories, prison has become a means to incapacitate rather than rehabilitate. Consequently, a new form of crime prevention infrastructure has emerged, which aims to create interconnected preventive partnerships. Current penal measures combine punitive justice with the

rhetoric of public protection. Research concerning inmates and drug abuse is vulnerable to be affected by political climate that reflects a dominant rhetoric of the time and place (Blagden and Pemberton 2010).

Garland suggests that this shift has introduced a new line of criminological theories. Unlike previous criminal theories that focused on individuals who commit crime, these new theories seek to predict and better manage criminal events and conditions that may induce crime and delinquency (Garland 2001). Simultaneously, a new form of crime prevention infrastructure has emerged, which aims to create local level preventive partnerships. The current model of crime control thus requires, and takes advantage of, surveillance as a technology of control, which is implemented by multiple related institutions. The result of the reinvention of the prison between 1973 and 1997 is reflected in the exponential increase in inmates incarcerated in the US (Schmitt et al. 2010). In contrast to the conventional wisdom of the previous period, the assumption now is that 'prison works' not as a mechanism of reform or rehabilitation, but as a means of incapacitation and punishment that satisfies popular political demands for public safety.

Just as political and social as the concept of crime itself, is the funding priority (Belenko 2002; Carnevale 1999; Meier 1992). Federal funding focus shifts from one project to another depending on political agenda, and interventions often are discontinued without plans for dissemination or routinized procedures folded into the correctional operation (Geneau et al. 2010; Hessels and van Lente 2008; Lee and Renzetti 1990; Rich 1989).

17.10 The Affordable Care Act and Changes in Insurance Coverage

Overall, 56% of state prisons and 33% of jails offer substance abuse treatment services (Taxman et al. 2007), and while more than 70% of state prisoners are estimated to be in need of drug treatment, only 13% actually receive care (Chandler et al. 2009); these proportions are even lower in jails (Peters et al. 2005; Rich et al. 2014). The National Institute on Drug Abuse (NIDA) reports that the services provided in correctional settings are often limited to behavioral interventions without medication therapy, which might be beneficial, and yet not comprehensive. NIDA has also argued that the division between drug abuse treatment and general healthcare makes it difficult to coordinate care and services for inmates who are often affected by multiple physical and mental health problems (Chandler et al. 2009; Mechanic 2012).

The Affordable Care Act (ACA) will hopefully secure more coordinated care for inmates with drug abuse and co-occurring conditions (Bainbridge 2012; Barry and Huskamp 2011; Phillips 2012), and thus have a significant effect on incarcerated populations, as an estimated 3.7 million uninsured individuals with mental illness are eligible for insurance coverage with the implementation of ACA (Miller 2014). The ACA's provision for behavioral health services is expected to expand substance abuse treatment programs (Buck 2011). As of today, expanded Medicaid coverage under the ACA expansion to those 138% of the poverty level would benefit many individuals with substance abuse problems.

Still, there are many questions. Close to 35% of newly insured individuals covered by the ACA's Medicaid expansion are known to have been involved with the criminal justice system (National Health Care for the Homeless Council 2013). However, the scope of Medicaid expansion varies by state. Falling below 138% of the federal poverty level income limit translates to an individual annual income of $11,880, or $24,300 for a family of four (HealthCare.Gov 2016). Many inmates, however, have low-wage, unstable jobs, whose income levels are in and out of the federal poverty line throughout a given year. Additionally, do drug treatment programs included in the ACA provisions require citizenship to be eligible for treatment? Such a question is especially relevant for the undocumented, who most likely utilize federally qualified health centers or other community clinics.

Along with the ACA implementation, research questions concerning inmates and substance abuse should be able to provide evidence

whether and/or how health care reform and expanded access to care affect health outcomes in the inmate population. Differences between states that have expanded Medicaid coverage and those that opted out are expected. One recent study reported that uninsured individuals are concentrated in states that have opted out of the Medicaid expansion (Shartzer et al. 2014).

Health disparities, in part, could be mitigated through access to quality care, and the implementation of the ACA is expected to improve drug treatment programs and other behavioral health services in civilian populations, which may, in turn, reduce incarceration and recidivism rates. However, it is also likely that without changes in social determinants and fundamental causes, disparities may persist, albeit through different mechanisms and pathways. To understand such structural factors and exposure to drug abuse and incarceration, neighborhood disadvantage and racial inequality become even more important research topics.

17.11 Implications: A Different Look

In recent years, the importance of linkages between related programs and datasets has been well documented in both correctional health and public health in general (Health Resources and Services Administration (HRSA) 2008; Substance Abuse and Mental Health Services Administration (SAMHSA) 2016). While discharge planning, re-entry, and seamless care are key to better outcomes, data linkage across institutions may contribute to unintended consequences in individuals with multiple health issues. Inmates are monitored and their activities are regulated at all time, which may increase adherence to treatment (Chandler et al. 2009). Research interventions often utilize monitoring and surveillance techniques for health problems and undesirable behaviors associated with these problems. The concept of linkage between corrections and community health care is based on the idea of continuity of care and treatment adherence, which, in many ways, require linking a myriad of data systems, including jails, hospitals, and drug and mental health treatment programs.

Connecting information from multiple entities can provide a more complete assessment of individuals who may potentially engage in future crime. In addition, to predict and preemptively manage crime, the criminal justice system needs to target high risk groups and locations by linking relevant institutions. For example, studies have documented that substance abuse, mental illness, and homelessness are often associated with crime and delinquency (Hartwell 2004); thus closely monitoring substance users, people with mental illness, and/or the homeless might seem to be an effective way to manage crime risk. In fact, many jails and prisons provide drug treatment programs to inmates, on the premise that these treatment programs are more effective with the continuity of services beyond correctional facilities. When inmates are released from corrections, they are often enrolled in community-based treatment programs (Kempf-Leonard 2000). Similarly, along with the disappearance of mental health facilities in the past decades, jails now provide more mental health services than any other community mental health facilities (Bullock 2011). Consequently, those who are considered high risk are under constant surveillance, and monitored by multiple institutions across a range of different settings (Ibarra et al. 2014). Power and social control has been always an issue understanding patient care provider relations in medicine (Goffman 1961; Scull 1977; Tiger 2011). The issue of power and control is even more prominent in the case of substance abuse. Inmates who need substance abuse treatment in particular are subject to such control because they are under constant monitoring/surveillance in the name of treatment (Dingel et al. 2011; Hunt and Barker 1999; Lyons 2014; McCorkel 2003; Timmermans and Gabe 2002).

Haggerty and Ericson (2000) argue that modern surveillance is not of control by separate, independently functioning institutions, but rather of assemblages of multiple processes that encompass a broad range of areas. The idea of "data doubles" is possible when multiple data

sources from different disciplines are merged into a coalescent system; social workers, healthcare providers, police, educators, and corrections feed their specialized knowledge into the risk profiling surveillance systems that control the body. Feeley and Simon describe that the new penology adopts techniques of managing and classifying risk groups, relying on surveillance and pre-emptive containment, which inevitably focuses more on probability and efficient risk control (Feeley and Simon 1992).

Haggerty and colleagues maintain that surveillance is a means to classify groups of people, which inevitably creates a hierarchical order among groups (Haggerty et al. 2011). Ibarra et al. (2014) argue that, although surveillance is ubiquitous in the modern world, the pervasive existence of surveillance, or "ambient surveillance", needs to be differentiated from the "interactive surveillance" that targets specific problem populations, thus requiring close monitoring and interventions. The authors refer to this focused surveillance as "casework" (Ibarra et al. 2014). Similarly, Lyon argues that data generated from surveillance and monitoring are used to establish categories and codes to classify individuals, which are then utilized for "social sorting" (Lyon 2008).

Thus, researchers need to be aware of how data can be used to further exclude and discriminate (former) inmates. Scholars have argued that incarceration is a means to exclude and contain those who are considered to be undesirable (Czajka 2005; Wacquant 2001), which, in essence, is a device of social control. Those who are condemned to be socially undesirable, as are those described in Goffman's *Asylum* and Foucault's *Discipline and Punish*, are under control and surveillance, keeping them from the general public (Timmermans and Gabe 2002). Foucault argued that healthcare providers are deeply engaged in separating, classifying, labeling, monitoring, and subjectifying those with undesirable symptoms and/or diseases (Foucault 1975, 1994). Substance abuse is labeled as a chronic disease and, at the same time, it is an illicit act. As such, the boundary between disease and crime becomes blurred, and at the border of these two seemingly contradicting conceptions of drug use, there is even more ambiguity. Substance abuse research in correctional facilities inevitably spans the border by explaining deviance and, at the same time, seeking effective monitoring and managing.

17.12 Conclusion

A growing body of literature has been devoted to understanding psychosocial and physical health concerns of incarcerated individuals. Yet, the complexity of intertwined behavioral, social, economic, and political factors shaping the patterns of drugs, crime, and incarceration is far from fully explained. The exponential growth in jail and prison populations in the past decades is attributed to the dramatic increase in drug-related incarceration (Drug Policy Alliance 2007; The National Center on Addition and Substance Abuse 2010). And many researchers and practitioners have made tremendous efforts to develop and provide effective drug treatment programs for incarcerated populations.

Different groups of study participants may interpret and rate items in questionnaires differently, and thus measurement invariance cannot be assumed when utilizing existing measurement tools, especially among vulnerable populations. In general, psychosocial and health measures are developed for the general population or the dominant group(s); and untested, these measurement tools may distort outcomes measured in subgroups. Studies have examined cross-cultural equivalence of many concepts and concluded that measurement tools need to be tested for compatibility in different groups. This holds true for within subgroups, such as racial/ethnic minorities or those who are excluded from mainstream institutions, including inmates. However, this issue of measurement equivalence is beyond the matter of refining measurement tools, rather, it may be an issue of reflexivity in the research process, especially for research concerning inmates who are often living in life

conditions that present multiple intersecting disadvantages. Perceptions and interpretations of inmates may be different from the general population, and need careful attention. However, researchers need to be reminded not to categorize inmates with simplistic binary categories; for instance, either mere victims of their circumstances or willful offenders. No mutually exclusive categories could capture complex conditions that affect inmates with substance problems, and ways in which they navigate and make sense of their circumstances.

In recent years, the field of intersecting issues of drug abuse and incarceration research has expanded its framework to include fundamental and structural causes of these problems. Many inmates experience a myriad of co-occurring physical and mental illnesses, in conjunction with substance abuse. High rates of mental health problems, increased risk for HIV, and other sexually transmitted diseases have been identified (Hammett 2006; Hammett et al. 1997; Maruschak 2006; Maruschak and Berzofsky 2015). Consequently, these co-occurring conditions need to be simultaneously dealt with in order to fully understand the magnitude of the burden. Furthermore, high prevalence of homelessness, poverty, low educational attainment, and joblessness has been documented among inmates and incarceration rates and other disadvantages tend to spatially cluster in highly segregated, poor minority neighborhoods.

These findings warrant a broader structural and contextual examination of substance users in corrections. For such research approaches, frameworks and analytic tools concerning drug abuse and incarceration should be able to explore multilevel, multifaceted, frequently inter-related causes. Perhaps it may call for different types of questions that go beyond traditional epidemiological approaches. In recent years, health disparities research and social epidemiology scholars have argued that we should move toward examining more distal, interconnected social determinants of diseases. These comprehensive approaches, such as the web of causation and ecological approaches, can be useful in research with the incarcerated population.

And yet, jails and prisons are difficult settings to actually implement such broader perspectives, in part due to the safety and security oriented, closed environment of correctional facilities. The inherent disconnect between corrections and the community makes it difficult to extend the causes and consequences of incarceration beyond jails and prisons. At the same time, corrections are often politically sensitive institutions that may be prone to sudden leadership change as a result of political turnover.

Even with a clear sense of purpose and dedication, conducting research in correctional settings is a daunting task. Although researchers fully understand the need for extra-heightened vigilance concerning research ethics, insufficiency and inexperience of IRB reviewers with the research topic may add unintended, additional burden to researchers. Researchers may need to consider ways to educate IRB reviewers, in terms of the importance of research concerning inmates and with means to prevent any adverse events.

Conducting research with inmates may present many challenges, but the task of understanding the effects of incarceration and drugs on individuals and communities is vital in eliminating health disparities and inequality. The role of researchers to contribute to knowledge production is not value-free. As Howard Becker once wrote, the question may be about whose side we are on. This does not mean that we have to be deeply affected by our own research, which we often are, and fail to deal with potential biases, which we can effectively avoid by being aware of the topics we discussed in this chapter. However, it may mean that we refuse to pretend that knowledge production is a completely apolitical process.

References

Abrams, L. (2010). Sampling 'hard to reach' populations in qualitative research: The case of incarcerated youth. *Qualitative Social Work, 9*(4), 536–550.

Adlera, N., Busha, N. R., & Pantell, M. S. (2012). Rigor, vigor, and the study of health disparities. *Proceedings of the National Academy of Sciences of the United States of America, 109*(Suppl 2), 17154–17159.

Alfred, M., & Chlup, D. (2009). Neoliberalism, illiteracy, and poverty: Framing the rise in black women's incarceration. *Western Journal of Black Studies, 33*(4), 240–249.

Bainbridge, A. (2012). *The affordable care act and criminal justice: Intersections and implications*. Washington, DC: U.S. Department of Justice.

Barry, C., & Huskamp, H. (2011). Moving beyond parity —Mental health and addiction care under the ACA. *New England Journal of Medicine, 365*(11), 973–975.

Becker, H. (1997). *Outsiders: Studies in the sociology of deviance*. New York, NY: The Free Press.

Belenko, S. (2002). The challenges of conducting research in drug treatment court settings. *Substance Use and Misuse, 37*(12&13), 1635–1664.

Binswanger, I., Krueger, P., & Steiner, J. (2009). Prevalence of chronic medical conditions among jail and prison inmates in the USA compared with the general population. *Journal of Epidemiology and Community Health, 63*(11), 912–919.

Binswanger, I., Nowels, C., Corsi, K., Glanz, J., Long, J., Booth, R., et al. (2012). Return to drug use and overdose after release from prison: A qualitative study of risk and protective factors. *Addiction Science & Clinical Practice, 7*(3), 1–9.

Binswanger, I., Stern, M., Deyo, R., Heagerty, P., Cheadle, A., Elmore, J., et al. (2007). Release from prison-a high risk of death for former inmates. *New England Journal of Medicine, 356*(2), 157–165.

Blagden, N., & Pemberton, S. (2010). The challenge in conducting qualitative research with convicted sex offenders. *The Howard Journal of Criminal Justice, 49*(3), 269–281.

Blandford, A. M., & Osher, F. (2013). *Guidelines for the successful transition of people with behavioral health disorders from jail and prison*. Delmar, NY: The Substance Abuse and Mental Health Services Administration's GAINS Center for Behavioral Health and Justice Transformation. Available at: http://gainscenter.samhsa.gov/

Boardman, J., Finch, B., Ellison, C., Williams, D., & Jackson, J. (2001). Neighborhood disadvantage, stress, and drug use among adults. *Journal of Health and Social Behavior, 42*(2), 151–165.

Bobo, L., & Thompson, V. (2010). Racialized mass incarceration: Poverty, prejudice, and punishment. In H. Markus & P. Moya (Eds.), *Doing race: 21 essays for the 21st century*. New York, NY: Norton.

Bollen, K. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53,* 605–634.

Bosworth, M., Campbell, D., Demby, B., Ferranti, S., & Santos, M. (2005). Doing prison research: Views from inside. *Qualitative Inquiry, 11*(2), 249–264.

Brewer, R., & Heitzeg, N. (2008). The racialization of crime and punishment: Criminal justice, color-blind racism, and the political economy of the prison industrial complex. *American Behavioral Scientist, 51*(5), 625–644.

Browning, C., Feinberg, S., & Dietz, R. (2004). The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. *Social Forces, 83*(2), 503–534.

Buck, J. (2011). The looming expansion and transformation of public substance abuse treatment under the affordable care act. *Health Affairs, 30*(8), 1402–1410.

Bullock, K. (2011). The construction and interpretation of risk management technologies in contemporary probation practice. *British Journal of Criminology, 51*(1), 120–125.

Bundy, C. (2004). Changing behaviour: Using motivational interviewing techniques. *Journal of the Royal Society of Medicine, 44*(97), 43–47.

Bursik, R., & Grasmick, H. (1993). *Neighborhoods and crime: The dimensions of effective community control*. New York, NY: Lexington Books.

Butzin, C. A., Martin, S. S., & Inciardi, J. A. (2002). Evaluating component effects of a prison-based treatment continuum. *Journal of Substance Abuse Treatment, 22*(2), 63–69.

Carnevale, J. (1999). Matching rhetoric to dollars: Twenty-five years of federal drug strategies and drug budgets. *Journal of Drug Issues, 29,* 299–321.

Carpenter, S. (2001). *Cognition is central to drug addiction*. Washington, DC: American Psychological Association. Available at: http://www.apa.org/monitor/jun01/cogcentral.aspx

Carson, E. A. (2015). *Prisoners in 2014*. Washington, DC: US Department of Justice (NCJ 248955).

Centers for Disease Control and Prevention. (2012). Estimated HIV incidence in the United States, 2007–2010. *HIV Surveillance Supplemental Report, 17*(4), 1–26.

Chandler, R. K., Fletcher, B. W., & Volkow, N. D. (2009). Treating drug abuse and addiction in the criminal justice system: Improving public health and safety. *Journal of the American Medical Association, 301*(2), 183–190.

Clear, T. R., Rose, D. R., Waring, E., & Scully, K. (2003). Coercive mobility and crime: A preliminary examination of concentrated incarceration and social disorganization. *Justice Quarterly, 20*(1), 33–64.

Community Oriented Correctional Health Services (COCHS). (2010). *Affiliations between health centers and local correctional facilities to provide continuity of care for offenders*. Oakland, CA: Community Oriented

Correctional Health Services (COCHS). Available at: http://www.cochs.org/affiliations-between-health-centers-and-local-correctional-facilities-provide-continuity-care-offend

Cook, R., Bernstein, A., & Andrews, C. (1997). Assessing drug use in the workplace: A comparison of self-report, urinalysis, and hair analysis. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates.* NIDA Research Monograph 167. Rockville, MD: US Department of Health and Human Services.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Cowburn, M. (2010). Principles, virtues and care: Ethical dilemmas in research with male sex offenders. *Psychology, Crime and Law, 16*(1–2), 65–74.

Coyle, M. (2002). *Race and class penalties in crack cocaine sentencing*. Washington, DC: The Sentencing Project. Available at: http://www.hawaii.edu/hivandaids/Race%20and%20Class%20Penalties%20in%20Crack%20Cocaine%20Sentencing.pdf

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Standford Law Review, 43*(6), 1241–1299.

Czajka, A. (2005). Inclusive exclusion: Citizenship and the American prisoner and prison. *Studies in Political Economy, 76,* 111–142.

Dankwa-Mullan, I., Rhee, K. B., Stoff, D. M., Pohlhaus, J. R., Sy, F. S., Stinson, N., et al. (2010). Moving toward paradigm-shifting research in health disparities through translational, transformational, and transdisciplinary approaches. *American Journal of Public Health, 100*(Suppl 1), S19–S24.

Dennis, M. L., Fetterman, D. M., & Sechrest, L. (1994). Integrating qualitative and quantitative evaluation methods in substance abuse research. *Evaluation and Program Planning, 17*(4), 419–427.

Dennis, M. L., Perl, H. I., Huebner, R. B., & McLellan, A. T. (2000). Twenty-five strategies for improving the design, implementation and analysis of health services research related to alcohol and other drug abuse treatment. *Addiction, 95*(Suppl 3), S281–S308.

Dingel, M. J., Karkazis, K., & Koenig, B. A. (2011). Framing nicotine addiction as a "disease of the brain": Social and ethical consequences. *Social Science Quarterly, 92*(5), 1363–1388.

Draine, J., Salzer, M. S., Culhane, D. P., & Hadley, T. (2002). Role of social disadvantage in crime, joblessness, and homelessness among persons with serious mental illness. *Psychiatric Services, 53*(5), 565–573.

Drug Policy Alliance. (2007). *Mandatory minimum sentences*. Available at: http://www.drugpolicy.org/drugwar/mandatorymin/

Du, J., Huang, D., Min, Z., & Hser, Y. I. (2013). Drug-abusing offenders with co-morbid mental disorders: Gender differences in problem severity, treatment participation, and recidivism. *Biomedical and Environmental Sciences, 26*(1), 32–39.

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2002). A multilevel analysis of neighborhood context and youth alcohol and drug problems. *Prevention Science, 3*(2), 125–133.

Duncan, T. E., Duncan, S. C., Okut, H., Strycker, L. A., & Hix-Small, H. (2003). A multilevel contextual model of neighborhood collective efficacy. *American Journal of Community Psychology, 32*(3/4), 245–252.

Evans, E., Longshore, D., Prendergast, M., & Urada, D. (2006). Evaluation of the substance abuse and crime prevention act: Client characteristics, treatment completion and re-offending three years after implementation. *Journal of Psychoactive Drugs, 38*(3 Suppl), 357–367.

Evans, G., & Kim, P. (2010). Multiple risk exposure as a potential explanatory mechanism for the socioeconomic status–health gradient. *Annals of the New York Academy of Science, 1186,* 174–189.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23,000 prisoners: A systematic review of 62 surveys. *The Lancet, 359*(9306), 545–550.

Federal Bureau of Prisons Clinical Practice Guidelines (FBOP). (2014). *Detoxification of chemically dependent inmates*. Federal Bureau of Prisons. Available at: http://www.bop.gov/resources/health_care_mngmt.jsp

Feeley, M., & Simon, J. (1992). The new penology. *Criminology, 30*(4), 449–474.

Fletcher, B., & Tims, F. (1992). Methodological issues: Drug abuse treatment research in prisons and jails. *NIDA Research Monograph Series, 118,* 246–260.

Ford, C., & Airhihenbuwa, C. (2010). The public health critical race methodology: Praxis for antiracism research. *Social Science and Medicine, 71*(8), 1390–1398.

Foster, H., & Hagan, J. (2007). Incarceration and intergenerational social exclusion. *Social Problems, 54*(4), 399–433.

Foucault, M. (1975). *Discipline and punish: The birth of the prison*. New York, NY: Vintage.

Foucault, M. (1994). Power. In J. D. Faubion (Ed.), *Essential works of Foucault 1954–1984* (Vol. 3). New York, NY: The New Press.

Fox, K., Zambrana, K., & Lane, J. (2011). Getting in (and staying in) when everyone else wants to get out: 10 lessons learned from conducting research with inmates. *Journal of Criminal Justice Education, 22*(2), 304–327.

Freudenberg, N. (2007). Health research behind bars: A brief guide to research in jails and prisons. In R. Greifinger (Ed.), *Public health behind bars*. New York, NY: Springer.

Furr, M., & Bacharach, V. (2008). Psychometrics and the importance of psychological measurement. In M. Furr & V. Bacharach (Eds.), *Psychometrics*. Thousand Oaks, CA: Sage Publications Inc.

Garland, D. (2001). *The culture of control*. Chicago, IL: University of Chicago Press.

Gelberg, L., Andersen, R. M., & Leake, B. D. (2000). The behavioral model for vulnerable populations:

Application to medical care use and outcomes for homeless people. *HSR: Health Services Research, 34* (6), 1273–1302.

Geneau, R., Stuckler, D., Stachenko, S., McKee, M., Ebrahim, S., Basu, S., et al. (2010). Raising the priority of preventing chronic diseases: A political process. *Lancet, 376*(13), 1689–1698.

Goffman, E. (1961). *Asylums: Essays on the social situation of mental patients and other inmates*. Garden City, NY: Anchor Books.

Goode, E., & Ben-Yehuda, N. (1994). Moral panics: Culture, politics, and social construction. *Annual Review of Sociology, 20,* 149–171.

Gostin, L. O., Vanchieri, C., & Pope, A. (Eds.). (2006). *Ethical considerations for research involving prisoners*. Washington, DC: Institute of Medicine, National Academies Press.

Greenberg, E., Dunleavy, E., & Kutner, M. (2007). *Literacy behind bars: Results from the 2003 National Assessment of adult literacy prison survey*. Washington, DC: U.S. Department of Education (NCES 2007–473).

Greenberg, G. A., & Rosenbheck, R. A. (2008). Jail incarceration, homelessness, and mental health: A national study. *Psychiatric Services, 59*(2), 170–177.

Grella, C. E., & Greenwell, L. (2007). Treatment needs and completion of community-based aftercare among substance-abusing women offenders. *Women's Health Issues, 17*(4), 244–255.

Haggerty, K., & Ericson, R. (2000). The surveillant assemblage. *British Journal of Sociology, 51*(4), 605–622.

Haggerty, K. D., Wilson, D., & Smith, G. J. D. (2011). Theorizing surveillance in crime control. *Theoretical Criminology, 15*(3), 231–237.

Hahn, E. A., & Cella, D. (2003). Health outcomes assessment in vulnerable populations: Measurement challenges and recommendations. *Archives of Physical Medicine and Rehabilitation, 84*(Suppl 2), S35–S42.

Hammett, T. (2006). HIV/AIDS and other infectious diseases among correctional inmates: Transmission, burden, and an appropriate response. *American Journal of Public Health, 96*(6), 974–978.

Hammett, T. M., Harmon, M. P., & Rhodes, W. (1997). The burden of infectious disease among inmates of and releasees from US correctional facilities, 1997. *American Journal of Public Health, 92*(11), 1789–1794.

Harkins, L., & Beech, A. (2007). Measurement of the effectiveness of sex offender treatment. *Aggression and Violent Behavior, 12,* 36–44.

Harrison, L. (1997). The validity of self-reported drug use in survey research: An overview and critique of research methods. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates*. NIDA Research Monograph 167. Rockville, MD: National Institute on Drug Abuse. Available at: https://archives.drugabuse.gov/pdf/monographs/monograph167/monograph167.pdf

Harrison, L., Martin, S., Enev, T., & Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population*. Rockville, MD: Substance Abuse and Mental Health Services Administration. (SMA 07-4249) Available at: https://www.cdhs.udel.edu/content-sub-site/Documents/Publications/Comparing%20Drug%20Testing%20and%20Self-Report%20of%20Drug%20Use%20Among%20Youths%20and%20Young%20Adults%20in%20the%20General%20Population.pdf

Hartwell, S. (2004). Triple stigma: Persons with mental illness and substance abuse problems in the criminal justice system. *Criminal Justice Policy Review, 15*(1), 84–99.

Health Resources and Services Administration (HRSA). (2008). *Enhancing linkages: Opening doors for jail inmates*. Washington, DC: Health Resources and Services Administration. Available at: http://hab.hrsa.gov/abouthab/files/cyberspns_linkages.pdf

HealthCare.Gov. (2016). *Federal poverty level (FPL)*. Baltimore, MD: US Centers for Medicare & Medicaid Services. Available at: https://www.healthcare.gov/glossary/federal-poverty-level-FPL/

Hessels, L. K., & van Lente, H. (2008). Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy, 37,* 740–760.

Homel, R., & Thomson, C. (2005). Causes and prevention of violence in prisons. In S. O. T. S. Eyland (Ed.), *Corrections criminology*. Sydney, AUS: Hawkins Press.

Hunt, G., & Barker, J. C. (1999). Drug treatment in contemporary anthropology and sociology. *European Addiction Research, 5,* 126–132.

Ibarra, P. P., Gur, O. M., & Erez, E. (2014). Surveillance as casework: Supervising domestic violence defendants with GPS technology. *Crime, Law and Social Change, 62*(4), 417–444.

Johnson, T. P., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005a). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 1–14.

Johnson, T. P., Cho, Y. I., Holbrook, A. L., O'Rourke, D., Warnecke, R. B., & Chavez, N. (2006). Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Annals of Epidemiology, 16,* 661–668.

Johnson, T. P., O'Rourke, D. P., Burris, J. E., & Warnecke, R. B. (2005b). An investigation of the effects of social desirability on the validity of self-reports of cancer screening behaviors. *Medical Care, 43*(6), 565–573.

Junger-Tas, J., & Marshall, I. H. (1999). The self-report methodology in crime research. *Crime and Justice, 25,* 291–367.

Kagawa-Singer, M. (2000). Improving the validity and generalizability of studies with underserved U.S. populations expanding the research paradigm. *Annals of Epidemiology, 10,* S92–S103.

Karberg, J., & James, D. (2005). *Substance dependence, abuse, and treatment of jail inmates, 2002*. Washington, DC: US Department of Justice. Available at: http://www.bjs.gov/content/pub/pdf/sdatji02.pdf

Kempf-Leonard, K. (2000). Expanding realms of the new penology: The advent of actuarial justice for juveniles. *Punishment and Society, 2*(1), 66–97.

King, R., & Mauer, M. (2002). *Distorted priorities: Drug offenders in state prisons*. Washington, DC: Sentencing Project. Available at: www.sentencingproject.org

Kitsuse, J., & Cicourel, A. (1963). A note on the uses of official statistics. *Social Problems, 11*(2), 131–139.

Kreft, I. G. G. (1994). Multilevel models for hierarchically nested data: Potential applications in substance abuse prevention research. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research*. NIDA Research Monograph 142. Rockville, MD: National Institute on Drug Abuse. Available at: https://archives.drugabuse.gov/pdf/monographs/142.pdf

Krieger, N., & Smith, G. D. (2004). Bodies count, and body counts: Social epidemiology and embodying inequality. *Epidemiologic Reviews, 26,* 92–103.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Auality & Quantity, 47*(4), 2025–2047.

La Vigne, N., Davies, E., Palmer, T., & Halberstadt, R. (2008). *Release planning for successful reentry*. Washington, DC: Urban Institute. Available at: http://www.urban.org/sites/default/files/alfresco/publication-pdfs/411767-Release-Planning-for-Successful-Reentry.PDF

Lee, R. M., & Renzetti, C. M. (1990). The problems of researching sensitive topics. *American Behavioral Scientist, 33,* 510–528.

Lennox, R. D., & Dennis, M. L. (1994). Measurement error issues in substance abuse services research: Lessons from structural equation modeling and psychometric theory. *Evaluation and Program Planning, 17*(4), 399–407.

Liebling, A. (2001). Whose side are we on?: Theory, practice and allegiances in prisons research. *British Journal of Criminology, 41,* 472–484.

Lipton, D. S. (1994). The correctional opportunity: Pathways to drug treatment for offenders. *Journal of Drug Issues, 24*(2), 331–348.

Lyon, D. (2008). *Surveillance society*. Available at: http://www.festivaldeldiritto.it/2008/pdf/interventi/divid_lyon.pdf

Lyons, T. (2014). Simultaneously treatable and punishable: Implications of the production of addicted subjects in a drug treatment court. *Addiction Research & Theory, 22*(4), 286–293.

Mackenzie, D. L. (2001). *Sentencing and corrections in the 21st century: Setting the stage for the future.* Rockville, MD: The National Criminal Justice Reference Service (NCJ 189089). Available at: https://www.ncjrs.gov/App/Publications/AlphaList.aspx?alpha=S&Agency=All

Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's sex offender treatment and evaluation program (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17,* 79–107.

Maruschak, L. (2006). *Medical problems of jail inmates*. Washington, DC: U.S. Department of Justice (NCJ 210696). Available at: http://www.bjs.gov/content/pub/pdf/mpji.pdf

Maruschak, L., & Berzofsky, M. (2015). *Medical Problems of State and Federal Prisoners and Jail Inmates, 2011–12*. Washington, DC: US Department of Justice. Available at: http://www.bjs.gov/content/pub/pdf/mpsfpji1112.pdf

Matheson, F. I., Doherty, S., & Grant, B. A. (2011). Community-based aftercare and return to custody in a national sample of substance-abusing women offenders. *American Journal of Public Health, 101*(6), 1126–1132.

Mauer, M. (2011). The challenges of implementing research based policies. *Criminology & Public Policy, 10*(1), 69–76.

Mauer, M., & King, R. (2007). *Uneven justice: State rates of incarceration by race and ethnicity*. Washington, DC: The Sentencing Project. Available at: http://www.sentencingproject.org/wp-content/uploads/2016/01/Uneven-Justice-State-Rates-of-Incarceration-by-Race-and-Ethnicity.pdf

McCorkel, J. A. (2003). Embodied surveillance and the gendering of punishment. *Journal of Contemporary Ethnography, 32,* 41–76.

McHugh, R. K., Hearon, B. A., & Otto, M. W. (2010). Cognitive-behavioral therapy for substance use disorders. *The Psychiatric Clinics of North America, 33*(3), 511–525.

Mechanic, D. (2012). Seizing opportunities under the affordable care act for transforming the mental and behavioral health system. *Health Affairs, 31*(2), 376–382.

Megargee, E. I. (1995). Assessment research in correctional settings: Methodological issues and practical problems. *Psychological Assessment, 7*(3), 359–366.

Meier, K. J. (1992). The politics of drug abuse: Laws, implementation, and consequences. *The Western Political Quarterly, 45*(1), 41–69.

Messina, N., & Grella, C. (2006). Childhood trauma and women's health outcomes in a California prison population. *American Journal of Public Health, 96* (10), 1842–1848.

Miller, J. (2014). *Dashed hopes, broken promises, more despair*. Alexandria, VA: American Mental Health Counselors Association (AMHCA). Available at: http://www.issuelab.org/resources/17415/17415.pdf

Miller, S. M., & Moulton, S. (2014). Publicness in policy environments: A multilevel analysis of substance abuse treatment services. *Journal of Public Administration Research and Theory, 24*(3), 553–589.

Minton, T., & Zeng, Z. (2015). *Jail inmates at midyear, 2014*. Washington, DC: US Department of Justice (NCJ 248629). Available at: http://www.bjs.gov/content/pub/pdf/jim14.pdf

Morenoff, J. D., & Sampson, R. J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces, 76,* 31–64.

Mumola, C., & Beck, A. (1997). *Prisoners in 1996*. Washington, DC: US Department of Justice (NCJ 164619). Available at: http://www.bjs.gov/content/pub/pdf/p.96.pdf

Mumola, C., & Karberg, J. (2006). *Drug use and dependence, state and federal prisoners 2004*. Washington, DC: US Department of Justice. Available at: http://www.bjs.gov/content/pub/pdf/dudsfp04.pdf

Muscat, B. (2008). *Violence and safety programs in women's prisons and jails: Addressing prevention, intervention and treatment*. Washington, DC: U.S. Department of Justice (225342). Available at: https://www.ncjrs.gov/pdffiles1/nij/grants/225342.pdf

National Health Care for the Homeless Council. (2013). *Medicaid expansion and criminal justice: Opportunities for the HCH community*. Nashville, TN: National Health Care for the Homeless Council. Available at: http://www.nhchc.org/wp-content/uploads/2011/10/NHCHC-MedicaidExpansion-Justice-Final.pdf

Newman, D. (1958). Research interviewing in prison. *Journal of Criminal Law and Criminology, 49*(2), 127–132.

Osher, F., Steadman, H. J., & Barr, H. (2003). A best practice approach to community reentry from jails for inmates with co-occurring disorders: The APIC model. *Crime & Delinquency, 49*(1), 79–96.

Paone, D., Chavkin, W., Willets, I., Friendmann, P., & Jarlais, D. D. (2009). The impact of sexual abuse: Implications for drug treatment. *Journal of Women's Health, 1*(2), 149–153.

Pelissier, B., Jones, N., & Cadigan, T. (2007). Drug treatment aftercare in the criminal justice system: A systematic review. *Journal of Substance Abuse Treatment, 32*(3), 311–320.

Peters, R. H., Greenbaum, P. E., Steinberg, M. L., Carter, C. R., Ortiz, M. M., Fry, B. C., et al. (2000). Effectiveness of screening instruments in detecting substance use disorders among prisoners. *Journal of Substance Abuse Treatment, 18,* 349–358.

Peters, R. H., Matthews, C. O., & Dvoskin, J. A. (2005). Treatment in prisons and jails. In J. H. Lowinson (Ed.), *Substance abuse: A comprehensive textbook*. Philadelphia, PA: Lippincott Williams & Wilkins.

Phillips, S. (2012). *The affordable care act: Implications for public safety and corrections populations*. Washington, DC: The Sentencing Project. Available at: https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=262818

Potter, J. (1996). *Representing reality: Discourse, rhetoric, and social constructionism*. London, UK: Sage.

Reimer, M., & Matthes, B. (2007). Collecting event histories with TrueTales: Techniques to improve autobiographical recall problems in standardized interviews. *Quality & Quantity, 41,* 711–735.

Renauer, B., Cunningham, W. S., Feyerherm, B., O'Connor, T., & Bellatty, P. (2006). Tipping the scales of justice. *Criminal Justice Policy Review, 17*(3), 362–379.

Rich, J. D., Chandler, R., Williams, B. A., Dumont, D., Wang, E. A., Taxman, F. S., et al. (2014). How health care reform can transform the health of criminal justice involved individuals. *Health Affairs, 33*(3), 462–467.

Rich, M. J. (1989). Distributive politics and the allocation of federal grants. *American Political Science Review, 83*(1), 193–213.

Richter, L., & Johnson, P. B. (2001). Current methods of assessing substance use: A review of strengths, problems, and developments. *Journal of Drug Issues, 31*(4), 809–832.

Rocque, M. (2011). Racial disparities in the criminal justice system and perceptions of legitimacy: A theoretical linkage. *Race and Justice, 1*(3), 292–315.

Rose, D., & Clear, T. (1998). Incarceration, social capital, and crime: Implications for social disorganization theory. *Criminology, 36*(3), 441–710.

Rose, D., & Clear, T. (2002). *Incarceration, reentry, and social capital: Social networks in the balance*. Paper prepared for the "From Prison to Home" Conference, Washington, DC (January, 2002). Available at: http://www.urban.org/UploadedPDF/410623_SocialCapital.pdf

Ross, C., & Mirowsky, J. (2001). Neighborhood disadvantage, disorder, and health. *Journal of Health and Social Behavior, 42*(3), 258–276.

Ross, C. E. (2000). Neighborhood disadvantage and adult depression. *Journal of Health and Social Behavior, 41* (2), 177–187.

Sampson, R. (2010). Punishment's place: The local concentration of mass incarceration. *Dædalus, 139* (3), 20–31.

Sampson, R., Morenoff, J., Smedley, B., & Syme, S. L. (2000). Public health and safety in context: Lessons from community-level theory on social capital. In B. D. Smedley & L. Syme (Eds.), *Promoting health: Intervention strategies from social and behavioral research*. Washington, DC: National Academy Press, Institute of Medicine.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94,* 774–802.

Sampson, R. J., & Jeghim-Bartusch, D. (1998). Legal cynicism and (subcultural?) tolerance of deviance: The neighborhood context of racial differences. *Law and Society Review, 32,* 777–804.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(15), 918–924.

Schlosser, J. A. (2008). Issues in interviewing inmates: Navigating the methodological landmines of prison research. *Qualitative Inquiry, 14*(8), 1500–1525.

Schmitt, J., Warner, K., & Gupta, S. (2010). *The high budgetary cost of incarceration*. Washington, DC: Center for Economic and Policy Research. Available at: http://cepr.net/documents/publications/incarceration-2010-06.pdf

Scull, A. T. (1977). Madness and segregative control: The rise of the insane asylum. *Social Problems, 24*(3), 337–351.

Sedlak, A., & McPherson, K. (2010). *Youth's needs and services: Findings from the survey of youth in residential placement*. Washington, DC: Department of Justice, Juvenile Justice Bulletin. Available at: https://www.ncjrs.gov/pdffiles1/ojjdp/227728.pdf

Sharkey, P., & Sampson, R. J. (2010). Destination effects: Residential mobility and trajectories of adolescent

violence in a stratified metropolis. *Criminology*, *48*(3), 639–681.

Shartzer, A., Kenney, G., Long, S., Hempstead, K., & Wissoker, D. (2014). *Who are the remaining uninsured as of June 2014?* Washington, DC: Urban Institute. Available at: http://www.urban.org/research/publication/who-are-remaining-uninsured-june-2014

Shaw, C., & McKay, H. (1942). *Juvenile delinquency and urban areas*. Chicago, IL: University of Chicago Press.

Sierles, F. (1984). Correlates of malingering. *Behavioral Sciences & the Law, 2,* 113–118.

Small, M. L. (2009). *Unanticipated gains*. New York, NY: Oxford University Press.

Small, M. L., Jacobs, E. M., & Massengill, R. P. (2008). Why organizational ties matter for neighborhood effects: Resource access through childcare centers. *Social Forces, 87*(1), 387–414.

Smedley, B. D., & Myers, H. F. (2014). Conceptual and methodological challenges for health disparities research and their policy implications. *Journal of Social Issues, 70*(2), 382–391.

Steiner, B. (2008). *Maintaining prison order: Understanding causes of inmate misconduct within and across Ohio correctional institutions*. Washington, DC: U.S. Department of Justice (226458). Available at: https://www.ncjrs.gov/pdffiles1/nij/grants/226458.pdf

Stevenson, B. (2011). *Drug policy, criminal justice and mass imprisonment*. Working paper prepared for the First Meeting of the Global Commission on Drug Policies. Geneva, CHE. Available at: http://www.globalcommissionondrugs.org/wp-content/themes/gcdp_v1/pdf/Global_Com_Bryan_Stevenson.pdf

Stewart, A., & Nápoles-Springer, A. (2000). Health-related quality-of-life assessments in diverse population groups in the United States. *Medical Care*, *38*(Suppl. II), II102–II124.

Stewart, A., & Nápoles-Springer, A. (2003). Advancing health disparities research: Can we afford to ignore measurement issues? *Medical Care, 41*(11), 1207–1220.

Stockdalea, S. E., Wells, K. B., Tang, L., Belin, T. R., Zhang, L., & Sherbourne, C. D. (2007). The importance of social context: Neighborhood stressors, stress-buffering mechanisms, and alcohol, drug, and mental health disorders. *Social Science and Medicine, 65*(9), 1867–1881.

Subramaniana, S. V., Kawachib, I., & Kennedy, B. P. (2001). Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Social Science and Medicine, 53*(1), 9–19.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2005). *Substance abuse treatment for adults in the criminal justice system*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at: http://store.samhsa.gov/product/TIP-44-Substance-Abuse-Treatment-for-Adults-in-the-Criminal-Justice-System/SMA13-4056

Substance Abuse and Mental Health Services Administration (SAMHSA). (2016). *SAMHSA's efforts on criminal and juvenile justice issues*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at: http://www.samhsa.gov/criminal-juvenile-justice/samhsas-efforts

Susser, M., & Susser, E. (1996). Choosing a future for epidemiology: II. From black box to Chinese boxes and ecoepidemiology. *American Journal of Public Health, 86*(5), 674–677.

Sutton, J. E., Bellair, P. E., Kowalski, B. R., Light, R., & Hutcherson, D. T. (2011). Reliability and validity of prisoner self-reports gathered using the life event calendar method. *Journal of Quantitative Criminology, 27*(2), 151–171.

Taxman, F. S., Perdoni, M. L., & Harrison, L. D. (2007). Drug treatment services for adult offenders: The state of the state. *Journal of Substance Abuse Treatment, 32*(3), 239–254.

The Centers for Disease Control and Prevention (CDC). (2001). *Helping inamtes return to the community*. Atlanta, GA: The U.S. Centers for Disease Control and Prevention. Available at: https://csgjusticecenter.org/nrrc/publications/helping-inmates-return-to-the-community-2/

The GAINS Center. (2004). *The prevalence of co-occurring mental illness and substance use disorders in jails*. Delmar, NY: The Substance Abuse and Mental Health Services Administration's GAINS Center for Behavioral Health and Justice Transformation, Fact Sheet Series.

The National Center on Addition and Substance Abuse. (2010). *Behind bars II, substance abuse and america's prison population*. New York, NY: Columbia University. Available at: http://www.centeronaddiction.org/addiction-research/reports/substance-abuse-prison-system-2010

The Sentencing Project. (2013). *Report of the sentencing project to the United Nations Human Rights Committee*. Washington, DC: The Sentencing Project. Available at: http://sentencingproject.org/doc/publications/rd_ICCPR%20Race%20and%20Justice%20Shadow%20Report.pdf

The Sentencing Project. (2014). *Trends in US corrections: US State and federal prison population, 1925–2014*. Washington, DC: The Sentencing Project. Available at: http://sentencingproject.org/doc/publications/inc_Trends_in_Corrections_Fact_sheet.pdf

Thomas, J. C., & Sampson, L. A. (2005). High rates of incarceration as a social force associated with community rates of sexually transmitted infection. *Journal of Infectious Diseases, 191*(Suppl 1), S55–S60.

Thomas, J. C., & Torrone, E. (2006). Incarceration as forced migration: Effects on selected community health outcomes. *American Journal of Public Health, 96*(10), 1762–1765.

Thomas, S. B., Quinn, S. C., Butler, J., Fryer, C. S., & Garza, M. A. (2011). Toward a fourth generation of disparities research to achieve health equity. *Annual Review of Public Health, 32,* 399–416.

Tiger, R. (2011). Drug courts and the logic of coerced treatment. *Sociological Forum, 26*(1), 169–182.

Timmermans, S., & Gabe, J. (2002). Introduction: Connecting criminology and sociology of health and illness. *Sociology of Health & Illness, 24*(5), 501–516.

Victor, J. S. (1998). Moral panics and the social construction of deviant behavior: Theory and application to the case of ritual child abuse. *Sociological Perspectives, 41,* 541–565.

Wacquant, L. (2001). The penalization of poverty and the rise of neo-liberalism. *European Journal on Criminal Policy and Research, 9,* 401–412.

Wacquant, L. (2009). *Punishing the poor*. Durham, NC: Duke University Press.

Wacquant, L. (2010). Class, race & hyperincarceration in revanchist America. *Daedalus, 139*(3), 74–90.

Wakefield, S., & Uggen, C. (2010). Incarceration and stratification. *Annual Review of Sociology, 36,* 387–406.

Wakeman, S., & Rich, J. (2010). HIV treatment in US prisons. *HIV Therapy, 4*(4), 505–510.

Webb, V. J., Katz, C. M., & Decker, S. H. (2006). Assessing the validity of self-reports by gang members: Results from the arrestee drug abuse monitoring program. *Crime Delinquency, 52*(2), 232–252.

Weiss, R. (1994). *Learning from strangers: The art and method of qualitative interview studies*. New York, NY: Simon & Schuster.

Western, B. (2006). *Punishment and inequality in America*. New York, NY: Russell Sage Foundation.

Western, B., Kling, J. R., & Weiman, D. F. (2001). The labor market consequences of incarceration. *Crime & Delinquency, 47*(3), 410–427.

Wilson, D. (2000). *Drug use, testing, and treatment in jails*. Washington, DC: Bureau of Justice Statistics (NCJ 179999).

Part VI
Policy Research

Application: What Role Does Research Play in Shaping Substance Abuse Policy?

18

John T. Carnevale

18.1 Introduction

In the United States, the federal government is recognized as the leader in addressing substance abuse and its consequences. In reality, the federal government does not act alone in formulating and implementing substance abuse policy. Because of our federalist system of government, state and local governments are also players in designing and delivering substance abuse polices—not necessarily in unison with the federal government. Together, federal, state, and local governments comprise the totality of governmental efforts at substance abuse control. In addition, the private sector promotes, develops, and implements substance abuse policies. Businesses may establish rules to regulate their workplaces to promote productivity. Foundations may sponsor prevention programs that they believe to accord with their definition of best practices. Advocacy groups promote policies and programs consistent with their specific goals and objectives. Volunteer organizations do the same. The international community, including the United Nation (with its numerous Conventions on drug policy[1]) and individual nations, will also represent their particular interests on substance abuse policy. The list of parties that are responsible for substance abuse policy seems endless, but the point is that there are many players in the policy formulation game. Each player brings their own views and beliefs—hopefully informed by research—to the design of substance abuse policy.

Research can be an important resource to help inform policymakers responsible for substance abuse policy. It is hoped that policymaker's goal is for substance abuse policy to be evidence-based,[2] meaning that it is informed by research and evaluation of effectiveness (maximization of desirable outcomes from a policy) and efficiency (implementing policies at the lowest per unit cost). Research can help policymakers achieve the goal of formulating and implementing evidence-based policies. Such research includes a vast array of activities ranging from sophisticated methods, discussed in this book, to simple analyses (e.g., data tabulations or graphical presentations) that an analyst may

[1]Many narcotic, plant-based, and psychotropic substances remain under international control under the 1961 Single Convention on Narcotic Drugs, the 1971 Convention on Psychotropic Substances, and the 1988 United Nations Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substance. The vast majority of governments are signatories to these international drug control treaties, which render the use, sale, traffic, and production of drugs like heroin, cocaine, and cannabis illegal.

[2]The term "evidence-based" is commonly used to denote policies that are informed by data and research. The phrase "evidence-based" has made its way into the parlance surrounding substance abuse or drug policy. Other similar phrases include "science-based" and "research-based." For purposes of this chapter, it is assumed that any of these descriptors of desirable policy may be used interchangeably.

J.T. Carnevale (✉)
Carnevale Associates, LLC, Gaithersburg, MD, USA
e-mail: John@carnevaleassociates.com

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_18

assert shows a finding important to substance abuse policy formulation, but in reality is viewed askance by those trained in research methods.

This chapter discusses the reality that all researchers must face: even the best and most relevant research may fail to inform policy. A host of factors contribute to this reality. First, policymakers can only use research that they are aware of—if the research does not reach the policymakers, they cannot use it. The clearance process offers one example of this barrier. Sometimes research is reported to a funder (e.g., a government agency or foundation) that has a sophisticated approval process before research is made available. In this case, it could be months before that research is cleared for release to policymakers. But even if research reaches policymakers, they may not understand it. Many policymakers are not trained in research methods and require a third party to provide translational services before research findings can be understood. Moreover, because research must be viewed within its proper context, policymakers may require contextualization from experts, even if they understand the specific research being considered. In many cases, contradictory evidence abounds; policymakers often require help understanding the nuances of research and distilling what (if any) policy implications exist. Ideology also comes into play, as policy exists with a political environment and few (if any) policymakers are free to consider research in a vacuum—that is, without some influence from their political and ideological framework. Finally, the policy change implied by research may be unfeasible—politically, bureaucratically, or for some other logistical reason. To adopt a policy change in this case might mean the loss of one's job, a challenge from program funders threatening to undermine an entity's resource base, or other similar serious concerns. For example, the popularity of supply reduction programs over demand reduction programs, as measured by federal funding for national drug control policies targeting illicit drugs, is in no small part due to a "tough on crime" stance by certain national policymakers (Reuter 2001).

In short, your research—no matter how relevant, timely, and significant—may not find its way to policymakers who design and shape substance abuse policies. Public policy may be shaped by the research that reaches policymakers—regardless of the quality of that research and the context within which it is placed for them. And decisions are made even when research cannot settle the question or even does not exist. Policymakers work with the research they have; and if they have nothing, they work with nothing. As will be discussed, much to the frustration of research scientists (Madras 2010; Reuter 2001), politics, ideology, media pressure, the public's desire for action, and bureaucratic needs forces decision-making in response to social or political pressures even when inaction may be subsequently determined to be preferable (Manski et al. 2001).

In this chapter, we will explore the reasons why research may not always find its way into the formulation of substance abuse policies. To do so, we will review the factors that are most prevalent in shaping substance abuse policy. The main theme is that research is but one ingredient of a substance abuse policy. In fact, there are many other ingredients in the mix that shape a policymaker's decision-making with regard to the design, structure, and implementation of a substance abuse policy. This book presents research in a normative—what ought to be—light as a rigorous tool used to ascertain information to inform and create evidence-based policy. This chapter shows that research can prevail in informing policy, but researchers must be realistic about their ability to influence policy development.

18.2 What Is a Substance Abuse Policy and Who Shapes It?

Substance abuse policy is defined here narrowly for purposes of simplicity to include policies involving psychoactive drugs or substances that are considered dangerous and may result in addiction. These include alcohol, tobacco, and

illicit drugs such as cocaine, heroin, and marijuana.[3] In the case of United States drug control policy, the Congress charged the Office of National Drug Control Policy (ONDCP) with developing a national drug control strategy to reduce the demand for and supply of illicit drugs.[4] Alcohol and tobacco are not included by law; however, ONDCP adopted the position that these two substances may be addressed in the national drug control strategy as part of national prevention efforts targeting underage (hence, illegal) use.

A local or state-level substance abuse policy may contain demand reduction components designed to reduce substance use and abuse through prevention, treatment, recovery, and support services. Sub-national policy may also contain elements to reduce supply through local law enforcement activities targeting illegal distribution, production, and cultivation of drugs as well as criminal behavior associated with drug use and supply. A national-level drug control policy may include these elements along with activities that are unique to a national government, including efforts to reduce the volume of drugs coming from outside a nation's borders—either by stopping (or interdicting) drugs at border crossings or eradicating drugs in their source countries. Even with this narrow view of a substance abuse policy, it should be apparent that the opportunities for research to inform it are vast. Many substance abuse policies exist at the national, state, and local levels at any point in time and the number of policymakers in both the public and private sectors who could benefit from knowledge derived from research is enormous.

We also define the policy development or formulation process as the process by which a substance abuse policy is created and formalized by policymakers. Given the potential number of substance abuse policies that the U.S. federal system could generate as well as the potential policies that may be formulated in the private sector, documentation of all these processes is impossible. Nevertheless, a desirable principle common to every one of them is that they embrace research to establish the best options for achieving desired outcomes. We introduce one policy formulation process in the next section that arguably represents a normative approach to policy formulation, where research is intended to play a substantial role.

18.3 A Policy Formulation Model

Strategic planning is a well-documented process that leverages knowledge about an underlying problem and a desire to mitigate the problem by identifying aims or desirable outcomes, addressing those aims through evidence-based programs and rigorously evaluating effectiveness. This section introduces a policy formulation process used by the Office of National Drug Control Policy in the late 1990s that is also a staple for substance abuse policy development by the Inter-American Drug Abuse Control Policy Commission (CICAD 2009).

The policy formulation process involves a systems approach that links four elements of a policy formulation process: (1) a needs assessment conducted by the community of stakeholders who have an interest in addressing substance abuse and its consequences; (2) a strategic plan that formalizes the parameters of a policy in terms of specific goals and objectives supported by measurable performance metrics; (3) evidence-based programs selected through the budget process to ensure proper implementation of the strategic plan; and (4) research and evaluation of policy and programs to assess performance relative to desirable results for

[3]As of the date of this writing (August 2016), marijuana has been legalized in four states in the United States (Alaska, Colorado, Oregon, and Washington) and the District of Columbia. Nevertheless, even though these states allow marijuana to be used for recreational purposes, it is classified as an illegal (schedule I) substance by the federal government under the Controlled Substance Act.

[4]The Office of National Drug Control Policy was established by the Anti-Drug Abuse Act of 1988 (P.L. 100-690) at the end of the Reagan Administration. It became operational in March 1989 during the beginning of the Bush Administration with the appointment of William Bennett as ONDCP's first Drug Czar.

process and outcomes (Simeone et al. 2005). This systems approach is integrative by design, whereby the needs assessment (also sometimes referred to as a situation assessment) drives policymakers to develop a policy and budget to implement that policy (a strategic plan and programs proposed through the budget formulation process), and a feedback mechanism to continuously assess performance of the plan and inform stakeholders through research and evaluation. This policy formulation process is dynamic. It presumes that policy will be changed over time and new information will be brought to bear on substance abuse policy. Figure 18.1 presents a simple depiction of this systems approach.

As seen in Fig. 18.1, "Community" refers to the stakeholders who have an interest in solving a particular drug problem. They gather in response to the problem, conduct a needs assessment, and determine the actionable areas that can be addressed by substance abuse policy. This plan is represented by the box labeled "Strategy." The strategy delineates the measurable goals and objectives to be pursued. It is an organizational tool that maps policy goals to actionable (and measurable) items. And it reflects the vision of the stakeholders and their specific efforts to reduce substance abuse and its negative consequences. The strategy serves as a plan or guide, designed to facilitate decision-making by program managers. The box labeled "Budget" refers to the resources necessary to implement the strategy. This usually reflects government resources, but it can also include nongovernment resources. Resources are intended to procure programs and activities that are evidence-based. In this policy formulation model, policy drives budget. The strategy informs those individuals responsible for developing and implementing an agency's budget about which programs to fund; it enables budget officers to justify budget formulation decisions. Finally, the box labeled "Evaluation" represents the feedback mechanism necessary to inform the stakeholders of their progress in achieving the goals and objectives explicitly delineated in the strategy. Evaluation uses data systems and research to assess strategic performance (efficiency and effectiveness), usually using identified performance targets and measures.

If utilized according to its normative design, knowledge from research and evaluation will be involved throughout the entire policy formulation process. For example, in conducting a needs assessment, data and policy analysis will be required to ascertain trends and patterns in substance abuse. This will likely entail the techniques discussed in this book, but one can assume that the identification of use and related consequences should employ the best research methodologies. With regard to implementation of the plan, where does the knowledge of evidence-based programs come from? In this ideal (perhaps utopian) policy formulation process, the answer is simple: research.

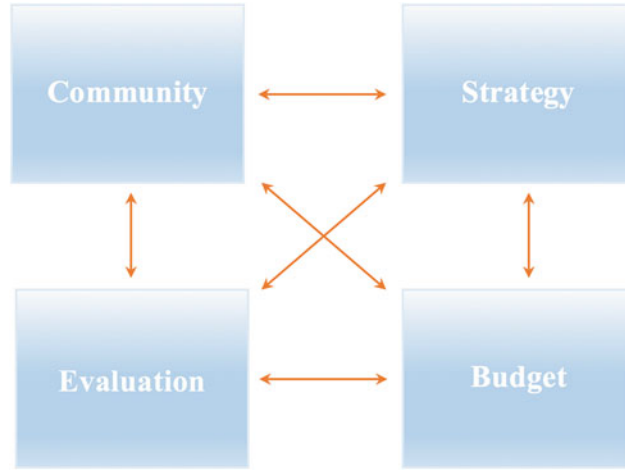

Fig. 18.1 An outcome-oriented strategic planning model

In reality, many in the substance abuse research community wonder why research has so little impact on policy (Reuter 2001). While the public sector embraces the use of evidence-based policies that require ongoing performance evaluation (Center for Substance Abuse Prevention 2009), cynics in the research community view this merely as rhetorical parlance used by policymakers to assert that their policy choices are evidence-based. Perhaps this cynicism is justified. We will soon see later in this chapter that sometimes—perhaps too often—research is overlooked, ignored, or dismissed for reasons unrelated to policy. These factors are the topic of the next section.

18.4 Determinants of Substance Abuse Policies

The previous discussion presents research as if it is the sole factor at work influencing substance abuse policy formulation. The truth is that research is not the sole factor shaping substance abuse policy—it is but one of many at work influencing a policymaker's decision-making. And research is not always without prejudice. As one scientist notes, the research community itself is suspect as there are questions about the "perceived legitimacy of science as an empirical means of finding the truth" and researchers themselves are not without bias (MacCoun 1998). Indeed, research is never perfect as it is a function of the reliability and validity of data systems it uses, the methodological approaches used to analyze data, and the potential biases within the research community who sometimes approach research with a specific agenda.

Toward the end of this chapter, we will review a case study in which researchers and their research fell in and out of favor in formulating one aspect of national drug control policy. The case study will show the interplay among a number of factors affecting policymaker's decision-making. But first, having considered research, we will now consider other factors that affecting the formulation of substance abuse policy.

Recall that in the substance abuse policy arena, policymakers are everywhere. They are in government or regulatory bodies responsible for making laws, rules, regulations, budgets, and so forth. They generally are accountable to the public, either directly as elected officials or indirectly as staff that report to elected officials (e.g., career staff in bureaucracies) or to taxpayers. Like everyone else, they gain knowledge and formulate opinions about issues from many sources besides research. Research has a role, sometimes a tremendous one, but its influence varies over time, by topic, and the influence of other factors affecting the formulation of policy. The following factors are the most salient ones in substance abuse policy. They are discussed as if they exist in isolation. But they are not mutually exclusive. In fact, they interact in many complex ways. They overlap; they sometimes reinforce one another, and they are sometimes at odds with one another. They include ideology, politics, public opinion, the media, bureaucracy, and advocacy.

As a final point, it may be worth mentioning that substance abuse policy is somewhat unusual because it is a policy area that most people feel they know something about. Many people believe that their personal interaction with the drug issue makes them an expert on the subject. Maybe this is because almost half of all Americans have tried an illicit drug in their lifetime or because over 8% of the population age 12 and older is classified with substance abuse dependence or abuse (SAMHSA 2014). Or maybe because the impact of substance abuse on society is far-reaching. For example, measured in terms of societal cost, the effects of illicit drug abuse are estimated to match those of other chronic health disorders. In fact, a recent study found that the $193 billion economic cost of drug abuse was on par with other serious health problems such as diabetes ($174 billion), obesity ($147 billion), and smoking ($157 billion) (NDIC 2011). Because of its breadth, most people know

someone with a substance abuse problem or have tried drugs themselves.

18.4.1 Ideology

In the public policy world, it is popular to claim that ideology—the system of beliefs, opinion, and ideas—is not an element of decision-making. But every person's views are shaped by their ideology (even if that ideology is also shaped by research). Even the best policymaker will have difficulty viewing an issue without ideological influences. And a cursory look at the day-to-day efforts of substance abuse policy demonstrates that ideology's policymaking role is significant. In the 1980s, when the nation was fed up with drug-related crime, a "tough on crime" ideology drove national decisions. As will be discussed below, many crime control laws were passed that brought a much more punitive approach to drug policy, the legacy of which mostly remains with us today (Reuter 2013).

More recently, national drug control policy released by the Obama Administration in 2009 emphasized a new approach—one that they claimed is based explicitly on science and not ideology (ONDCP 2010). By asserting this, the Administration was attempting to distinguish its drug control strategy from those of the previous four Administrations (Bush, Clinton, Bush, and Reagan), pointing out that it was unique because it was based on data and research and not personal beliefs. The truth of that assertion about a new approach will ultimately be a question for historians. But as one of the strategy's basic tenets, it resonated with individuals who were fed up with the "war on drugs" language used to describe drug policy since the Nixon Administration. Nevertheless, the Administration's rejection of ideology as a driver of drug policy implied that its policy was research-based. As evidence of its assertion, the 2010 Strategy used the substantial body of scientific evidence about addiction being a disease of the brain and promoted a healthcare approach to treating addiction. This is a marked rhetorical change compared to ideology-driven drug strategies constructed around the belief that addiction was a moral failure.

Ideology also shapes research, in some cases driving *what* research is initiated and very often also shaping how research is selected, interpreted, and understood. For example, foundations and "think tanks" have political, social, economic, or other agendas that define their ideological perspective; so do universities. Ideologies influence the types of research that these organizations fund and conduct. In the area of substance use policy, policymakers know which think tanks support their ideology and political leanings, and they may seek out research from those organizations that have similar ideological perspectives or reject research from competing organizations. For example, some policymakers might seek out the results of a study from the CATO Institute, which finds states that allow medical marijuana have fewer suicides than states that do not (Anderson et al. 2015). In contrast, other policymakers might seek out recent research showing that rescheduling marijuana from Schedule I (having no medical value) to Schedule II is premature because the current body of literature does not support marijuana's medical value (Murray 2014). Adherence to a specific ideological perspective by policymakers may also obviate the need for research on certain topics. For example, policymakers that believe the illegality of drug use is sufficient to justify enforcement may believe that research on the effectiveness of that enforcement is unnecessary (Reuter 2001). This discussion demonstrates that, despite claims to the contrary, ideology's reach in policy formulation must be recognized for what it is—an important determinant of substance abuse policy.

18.4.2 Politics

Politics is another factor in shaping substance abuse policy. Broadly speaking, there is bipartisan support for a national drug policy that seeks to reduce drug use and its consequences. Republicans and Democrats agree that the key ingredients of national policy should include law enforcement, substance abuse prevention and

treatment, interdiction aimed at preventing illicit drugs from entering the nation, and international programs designed to eliminate drugs at their source of cultivation or production (Carnevale Associates 2008). While there is general agreement on the basic ingredients of a national substance abuse policy, there has been substantial disagreement on how best to mix them to achieve the objective of reducing use and consequences. Republicans are traditionally associated with a "tough on crime" approach that emphasizes law enforcement, interdiction, and source country programs over treatment prevention (Meier 1990). Democrats are traditionally associated with a healthcare-oriented approach that emphasizes treatment and prevention over enforcement and interdiction. According to the one poll, the public's view about the merit of legalization of marijuana is divided across partisan line: Republicans continue to be far less likely than Democrats to favor legalization (39% vs. 63%) (Pew 2014).

Interestingly, while each political party has had its moments leading the formulation of drug policy since passage of the original legislation creating the Office of National Drug Control Policy in 1988, the difference in political views did not result in a significant change in the federal drug control budget—either the budget proposed by the Administration or the budget actually funded by the Congress (Carnevale and Murphy 1999; Reuter 2013.). However, while the budget for federal agency programs shows little sensitivity to the politics of drug control, politics is most manifest in laws and regulations.

Between 1984 and 1994, national laws targeting drug-related crime and strengthening law enforcement were more prevalent during Republican administrations than during Democratic ones. During this period, the Congress enacted five major anti-crime bills (Congressional Research Service 2007).[5] Four bills became law during the Reagan and Bush Administrations and the last one during the first term of the Clinton Administration.[6] No major crime control legislation was passed since then. The next significant national legislation affecting drug policy occurred during the Obama Administration when the Mental Health Parity and Addiction Equity Act of 2008[7] built upon the Mental Health Parity Act of 1996 (originally passed in the Clinton Administration). Later the Affordable Care Act of 2010 (ACA) would make even more significant changes to substance abuse treatment and prevention.[8] Among other things, the ACA was meant to support the public health approach rhetoric of the 2010 National Drug Control Strategy.

Political feasibility—the realistic chance of making a policy change weighed against the political (and literal) capital required to bring that change about—also has a tremendous influence of policy decisions. Such feasibility is necessarily an amalgam of the factors discussed in this chapter; politics, ideology, public opinion, and other factors all come together to limit policy options—even in the face of research. But because politics attempts to consider many of these other factors, political feasibility is addressed in this section.

Consider the example of state-level marijuana legalization. To date, research can only settle a handful of the questions needed to construct truly evidence-based policies, and legitimate debate remains about the competing goals of legalization (Calkins et al. 2016). However, the research community abounds with versions of marijuana legalization that are believed to be better for the public health than the for profit "alcohol style" distribution model that has so far dominated the industry, including "for benefit" corporations, co-op distribution and government monopolies, to name but a few (Caulkins et al. 2015). Yet,

[5]These include the following laws: The Crime Control Act of 1984 (P.L. 98-473); the Anti-Drug Abuse Act of 1986 (P.L. 99-570); the Anti-Drug Abuse Act of 1988 (P.L. 100-690); the Crime Control Act of 1990 (P.L. 101-647); and the Violent Crime Control and Law Enforcement Act of 1994 (P.L. 103-322).

[6]The Violent Crime Control and Law Enforcement Act of 1994 was promoted during the Clinton Administration.

[7]P.L. 110-343.

[8]The Affordable Care Act actually refers to two separate pieces of legislation—the Patient Protection and Affordable Care Act of 2010 (P.L. 111-148) and the Health Care and Education Reconciliation Act of 2010 (P.L. 111-152).

these policy solutions seem unlikely to pass in any states currently considering legalization. Why is that?

First, recall that no policy exists in a vacuum. In a world of finite resources, unrelated policies are in competition—for attention and for dollars. Time and money spent on one issue cannot be spent on another. A political party that chooses to prioritize substance abuse policy makes a choice to prioritize that issue over other issues. Seemingly, few parties appear willing to prioritize marijuana legalization in this way.[9] Second, note that novel approaches to any policy require the most change and consequently meet the most resistance. Legalizing on an alcohol distribution model allows policymakers and industry players to build off of existing systems. Creating a new regulatory system without an example on which to base it is a more difficult task. Third, note that the media and public opinion play a large role. Public opinion polls monitor opinions "for" and "against" legalization, but those decisions are largely un-nuanced. Media reports, too, often lack the nuance needed to convey major policy differences. Ideology is also a key factor: free market commercial consumption and sale is arguably a core value of many Americans, spanning political parties. The case of state-level marijuana legalization offers an example of how nearly every factor that influences policy can trump research, particularly when dealing with the details of implementation.

18.4.3 Public Opinion

Another factor influencing policymaker's decisions about substance abuse policy is public opinion, most often illuminated by polls conducted with some frequency (monthly, quarterly, annually). Many policymakers are politicians, and still more policymakers are supervised by political appointees. In a political system where policymakers are accountable to the public, public opinion matters. So public opinion—whether well-informed or ill-informed—plays a large role in policymaking. In the face of uncertainty, adopting a course of action that accords with public opinion may be less risky to job security than following research that runs contrary to public opinion. For example, if crime rates are high and the public thinks that drug use and crime are inextricably linked, then policymakers may respond by adopting policies targeting drug-related crime. During the 1980s, the public's view that drugs were the leading cause of crime increased from 13% in 1981 to 60% in 1990 (Roper Center for Public Opinion Research). At the same time, news reports linking drug use to crime increased significantly (Blendon 1998). It was during this period that policymakers adopted most of the crime control laws that remain in effect today. Passage of these laws occurred certainly during the period when public opinion about the drugs/crime nexus was most prominent.[10]

Importantly, public opinion and politics (as well as the media) are inexorably linked in a democratic system. Because politicians rely on the public for both their jobs and their governing authority, they also look to the public to guide their actions—perhaps even when they should be attempting to educate the public about research that contradicts public opinion. The path of least resistance for a politician (or a government employee that is ultimately accountable to a politician) will always be to favor public opinion over research.

18.4.4 The Media

The media is another factor that shapes substance abuse policy. The relationship between public opinion and the media is complex and

[9] This is especially noticeable as marijuana legalization increasingly passes by ballot initiative rather than through legislative action by elected officials.

[10] As a further example of the strength of public opinion in influencing policymakers, during his inaugural address in January 1989, George H.W. Bush said "There are few clear lines in which we as a society must rise up united and express our intolerance. The most obvious now is drugs… there is much to be done and to be said, but take my word for it: This scourge will stop." In March 1989, William Bennett was appointed to become the first drug czar in the Office of National Drug Control Policy.

bidirectional. The media affects what information is disseminated to the public (as well as how that information is disseminated and in what context). In that way, the media affects public opinion. However, public opinion also affects the media—as media outlets will make coverage decisions based on what they think will capture the interest of their audience. Both of these factors together, in turn, influence policymakers. So the media has the ability to systematically promote a message and can also report individual news events in real time, each one of which may not be part of an overarching message. The media also represents a tremendous number of avenues through which information reaches policymakers. It includes radio, television, news publications, and standard Internet outlets. More recently, social media's role is also increasingly relevant—perhaps reflecting a more "raw" version of public opinion (e.g., trends on Twitter) but also curated by the social media platforms themselves (e.g., Facebook's algorithm determining what articles to display on an individual's "news feed").

Regardless of the source, one has to only hearken back to a quote from Thomas Jefferson, "The man who reads nothing at all is better educated than the man who reads nothing but newspapers." Whatever one's media of choice, this quote is as entertaining today as it was back in Jefferson's time. While the factual accuracy of media reports may be questioned, there is no doubt about the power of their reach. For example, one study looked at the relationship in the 1980s between the influence of public opinion on drug policy and media stories and supported research finding that issues that occurred the most in the media were the issues that the public believed to be the most important (Bare 1990). It then went on to assert that the media was able to characterize the substance abuse problem as the most significant problem of the day, which in turn further affected politics and ideology.

18.4.5 Bureaucracy

The implementation of substance abuse policy requires action by the bureaucratic structure involved in delivering programs to address substance abuse and its consequences. Designated here as simply "bureaucracy," the term is intended to identify agencies, bureaus, departments, and so forth that share the characteristic of having a separate and distinct budget to finance their activities and are considered regularly in the government's budget formulation process. Of course, the term bureaucracy also includes any private sector entity such as a foundation or independent organization involved in addressing substance control matters.

Bureaucracies have constituencies and an interest in influencing those who finance their operations. The executive leadership in any bureaucracy has an interest in targeting their efforts to influence the political process that supports their operations by influencing policymakers (Meier 1990). Despite efforts to measure the success of bureaucracies through performance accountability, the reality is that the chief executive of any bureaucracy will likely measure his or her success by growth in his or her appropriations. A growing budget will signal policymaker confidence in a bureaucracy's operations. So, why is this information important to understanding the effects of research on substance abuse policy? The answer is simple: bureaucracies will promote policies and budgets to support their own self-interest, often even when research finds such action counterintuitive. Thinking back to the discussion about the mix of program ingredients that go into a national drug control policy, while the national policy direction may dictate a reduction in some ingredients (say, drug interdiction) and increases in others (say, prevention) based on research, the fact is that each of those areas is controlled by a distinct bureaucracy or bureaucracies. And individual bureaucracies will likely aggressively act to protect their resource base. Managing with more resources is easier than managing with less.

Competition over finite resources is not the only way by which bureaucracy influences policy to the detriment of research. Bureaucracy is resistant to major change—so any major policy change (even if supported by research) will encounter correspondingly major resistance if

implementing it would require significant infrastructure modification or significant changes to the status quo. Consider state probation and parole systems. They are frequently considered areas of drug control policy that require revision or reexamination, yet even if research suggested a dramatic shift in how a state runs its probation system, such a change is unlikely to come to fruition. Public opinion, ideology, and politics all play a role—but bureaucratic inertia may well play the largest role. Changing how a branch of government performs a core function requires a paradigm shift. History teaches that, without major legislative change, bureaucracies do not attempt such policy changes, even if legislative change is not theoretically required.

18.4.6 Advocacy

Finally, so long as limited resources compete for unlimited wants, advocacy will have a role in promoting policies, programs, and budgets. Advocacy is defined here as a process of organizing information in a way that most effectively promotes a preferred outcome. Advocacy is usually associated with lobbyists who may be viewed as paid advocates for a cause (e.g., prevention, a treatment program, insurance coverage) or organizations that receive appropriations. Advocacy is viewed as legitimate when it occurs in an "explicitly advocacy-based organization, or an explicitly adversarial system of disputing" (MacCoun 1998). While it can help inform the policymaking process by enlightening policymakers about the latest trends and patterns and research surrounding a topic, it can also bias the decision-making process through the selective use of information to cast the best light on a particular problem or solution. To be more direct, advocacy can be a fraudulent process if it results in concealment or distortion of evidence available to inform a topic. This is not to say that advocacy is bad; it is an important tool for disseminating information to inform the policy process so long as it is fact-based and not driven by ideology. As is the case with the other factors discussed, advocacy has both positive and negative aspects.

18.5 Case Study: Dueling Research Confronts Politics and Ideology

In this section, we review one example substance abuse policy in which research plays a substantial role in moving drug policy in one direction and then the opposite direction, only to be eventually determined by an independent body to be unreliable. In the end, the research was dismissed thereby leaving other factors—such as bureaucratic forces, politics, and ideology—to take charge.

The case study concerns a large program area in U.S. national drug control policy: drug interdiction. It shows how researchers provided contradictory evidence at different points during the 1990s about the cost-effectiveness of interdiction and how this research shaped interdiction policy. It also looks at how research collided head on with politics and ideology and how that collision affected policy. Finally, the case study shows how an independent scientific body—the National Academy of Science—decided that none of the research on interdiction was or could be valid, thereby leaving policymakers to fend for themselves.

Drug interdiction is defined as activities intended to reduce the availability of illegal drugs entering the United States by targeting the transportation link.[11] In other words, interdiction is about law enforcement and military efforts to prevent drugs from entering the United States. The most prominent federal agencies involved in drug interdiction are the U.S. Customs Service, the Border Patrol, the Coast Guard, and the Department of Defense. Before, these agencies were separate and distinct; now, all but the Department of Defense are part of the newly formed Department of Homeland Security.

Interdiction found prominence in U.S. drug control policy starting in the 1980s when the nation was focused on the cocaine epidemic. Many policymakers believed that preventing cocaine from entering the country would end the epidemic and the criminal activity associated

[11]ONDCP Circular: Budget Formulation, 2013. http://www.whitehouse.gov/sites/default/files/docs/2013_circular-budget_formulation.pdf.

with transportation and distribution. This thinking seemed logical to them, as coca, the plant from which cocaine is derived, is not endemic to the U.S. and comes from a handful of Andean nations. This thinking had ideological appeal, as it meant that other nations could be blamed for the United States' drug problem.

Federal spending on drug interdiction grew enormously during the 1980s on the belief that it would solve the cocaine problem. As a share of total federal spending, interdiction reached about 40% of the total drug control budget (Carnevale and Murphy 1999). Studies about the cost-effectiveness of interdiction did not first appear until near the end of the decade.[12] In 1988, a RAND study reporting that interdiction was relatively ineffective in stopping the flow of drugs or affecting prices (Reuter et al. 1988) was largely ignored. When the Office of National Drug Control Policy was formed in 1989, the strategies that it promoted during the Bush Administration continued to push for more interdiction spending. Supporters of interdiction viewed studies of law enforcement as unnecessary. Law enforcement was just that: efforts to reduce crime and increase public safety. Efforts to reduce enforcement meant being soft on crime, and no policymaker wanted to own that label.

When Bill Clinton assumed the presidency, a RAND Corporation study prepared for the Office of National Drug Control Policy looked at the relative cost-effectiveness of interdiction compared to demand control (Rydell and Everingham 1994). It found that a marginal dollar spent on drug control would have the greatest return on investment if spent on drug treatment. Interdiction had the lowest rate of return. In other words, interdiction was the least cost-effective program area. Coupled with another RAND study that found that interdiction would not notably raise drug prices or reduce availability (Caulkins et al. 1993), the two RAND studies had a profound impact on policymakers within the new Clinton Administration. Drug budgets proposed during Clinton's first term requested more resources for treatment and significantly less resources for interdiction (and programs targeting cultivation in source countries). While research prevailed in convincing the Clinton Administration to change course on interdiction, it did not persuade Congress. Appropriation committees rejected the increased spending requested for treatment, but reduced spending on interdiction only modestly. In this case, politics and ideology prevailed over research.

While not markedly affecting spending on interdiction programs, these research studies launched an intense debate about the cost-effectiveness of interdiction during the Clinton Administration's second term. A new drug czar—a four-star general appointed by the President—had to go toe-to-toe with Congressional appropriators who were convinced that the two RAND studies were flawed. Their case was bolstered by yet another study, this one from the Institute of Defense Analysis (IDA), that found interdiction to be cost-effective (Crane et al. 1997). The release of this IDA study that contradicted the RAND studies led to a Congressional oversight hearing designed to prove that the Clinton Administration's views about the relative ineffectiveness of interdiction were wrong. Politics and ideology were center stage in a debate over dueling research findings. The hearing failed to achieve its goal because of a Letter to the Editor from the Office of National Drug Control Policy on the day of the hearing that reported that the IDA study had not only not been cleared by that Institute, but that IDA's own leadership had doubts about the reliability of the findings. Albeit a petty political ploy, the letter to the editor cast doubt about validity of the IDA study. In this case, the media also affected policy.

But the story does not end there. Because of the importance of the issue to the appropriations process, bureaucratic pressure to increase budgets beyond what the drug czar requested, politics and ideology came together at a subsequent hearing where the oversight committee produced its own research on the topic. While not rigorous by classical research science standards, what constituted research was a graph (recreated in Fig. 18.2) showing the relationship between interdiction spending over time and youth drug

[12]This is generally consistent with the "research lag" that plagues even the most research-focused policymaker.

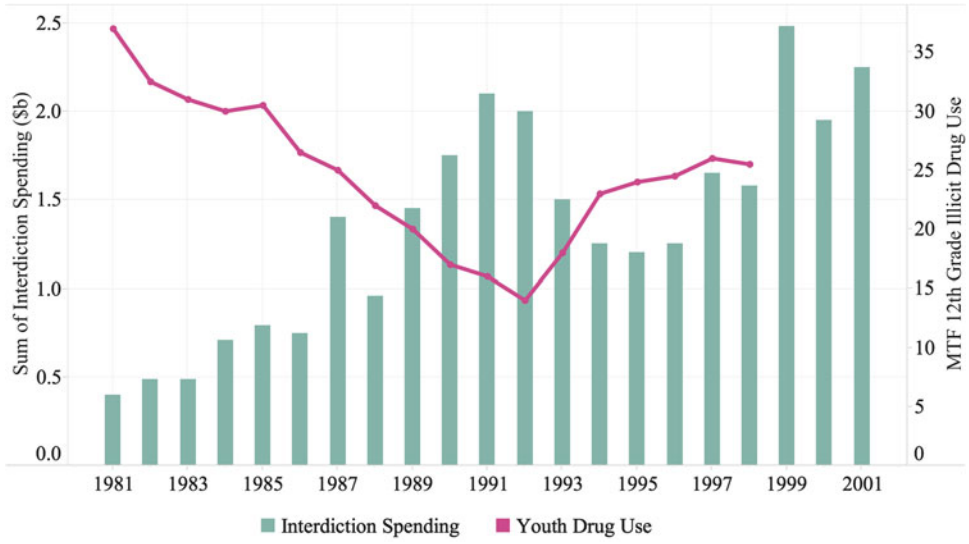

Fig. 18.2 Interdiction spending versus youth drug use trends

use. As the chart showed, interdiction spending and youth drug use were inversely related; youth drug use had been increasing throughout the 1990s when interdiction spending was reduced. The Committee made the case that the observed inverse relation depicted in the chart was positive evidence that interdiction was effective at reducing drug use. The solution to the nation's growing youth drug use problem? Increase interdiction spending. The Administration's drug czar did testify that what the committee was depicting appeared to be a classic case of spurious correlation, as youth drug use depicted in the chart represented mostly marijuana which was not the target of interdiction efforts. In fact, drug interdiction spending was allocated for federal agency activities targeting cocaine and heroin being shipped from South America to the United States. Nevertheless, despite the questionable reliability of the Committee's research, politics and ideology prevailed. Spending on drug interdiction was slowly restored over the remainder of the decade.[13]

In the wake of the controversy created by the conflict between research, politics, ideology, and to some extent bureaucratic pressures to sustain budgets, in early 1998, the Office of National Drug Control Policy sought to settle the issue about the efficacy of the RAND and IDA studies to inform the policy process. It funded an independent study by the National Academy of Science's National Research Council (NRC) to study data and research needed to inform drug policy, including an in depth review of the RAND and IDA studies. After reviewing the RAND and IDA studies, the NRC found that they did not produce plausible findings (Manski et al. 1999). The NRC reported that RAND did not provide "usable empirical findings on the relative cost-effectiveness of alternative polices in reducing cocaine consumption" and that the IDA study did "not yield useful empirical findings on the cost-effectiveness of interdiction policies to reduce cocaine consumption."[14] In its

[13]Interdiction spending increased by $1.1 billion from $1.1 billion in FY 1994 to $2.2 billion by FY 1999.

[14]Text reported in the Executive Summary of the 1999 NRC Report. The Executive Summary further stated that: "The [Rand] study makes many unsubstantiated assumptions about the processes through which cocaine is produced, distributed, and consumed. Plausible changes in these assumptions can change not only the quantitative findings reported, but also the main qualitative conclusions of the study. Hence the study's findings do not constitute a persuasive basis for the formation of cocaine control policy"; and that "major concerns about data and methods make it impossible to accept the IDA findings as a basis for the assessment of interdiction policies."

final report on the subject of data and research needed to inform drug policy, the NRC committee concluded that data shortcomings related to measuring drug use (data counts users rather than measure consumption) and drug prices (price data are collected by the Drug Enforcement Administration for purposes of providing evidence in criminal trial rather than obtaining market price information) were inadequate for measuring the cost-effectiveness of enforcement. So, not only did the NRC committee dismiss the RAND and IDA studies, its critique about the inadequacy of data systems meant that any attempts to measure the cost-effectiveness of enforcing interdiction policies would fail. Stated perhaps more directly, research into the effectiveness of interdiction would be of no value as a factor in informing drug interdiction policy. Hence, the U.S. would have to rely on the other factors to set its drug policy until the data gaps were closed.

In the end, the NRC's findings about data and methodological deficiencies had the result of ending all formal research into the topic. Fourteen years later, interdiction continues to play a large role in U.S. drug control policy even though knowledge of its cost-effectiveness or contribution to reducing availability, changing prices, and demand remains unknown. Until such time as improvements in data occur, ideology, politics, and bureaucratic interests will have to suffice in informing U.S. substance abuse policy on this particular topic.

18.6 Conclusion

Research can and does shape substance abuse policy. However, it is but one of many factors influencing the decisions about how to allocate limited resources among unlimited wants. Other factors are at work that sometimes complement or compete for favor in promoting what is rhetorically described as evidence-based policies. In the substance abuse policy arena, evidence from sophisticated research is often fleeting or nonexistent, but the need to make policy choices is not. Policymakers must make their decisions even when uncertainty abounds about results, both intended and unintended.

References

Anderson, D., Rees, D., & Sabia, J. (2015). High on life? Medical marijuana laws and suicide. *CATO Institute, Research Briefs in Economic Policy*, No 17.

Bare, J. (1990). The war on drugs: A case study in opinion formation. *The Public Perspective*, November/December.

Blendon, R., & Young, J. (1998). The public and the war on illicit drugs. *Journal of the American Medical Association, 279*(11), 827–832.

Carnevale Associates. (2008). *Fixing national drug control policy: Principles of an effective drug control policy*. Carnevale Associates Policy Brief. Available at: http://www.carnevaleassociates.com/fixing_drug_policy2008

Carnevale, J., & Murphy, P. (1999). Matching rhetoric to dollars: Twenty-five years of federal drug strategies and budgets. *Journal of Drug Issues, 29,* 299–322.

Caulkins, J., Crawford, G., & Reuter, P. (1993). Simulation of adaptive response: A model of interdictor-smuggler interactions. *Computer and Mathematical Modeling, 17*(2), 37–52.

Caulkins, J., Kilmer, B., & Kleiman, M. (2016). *Marijuana legalization: What everyone needs to know* (2nd ed.). Oxford, UK: Oxford University Press.

Caulkins, J., Kilmer, B., Kleiman, M., MacCoun, R., Midgette, G., Oglesby, P., et al. (2015). *The marijuana legalization Debate: Insights from Vermont*. Santa Monica, CA: RAND Corporation.

Center for Substance Abuse Prevention. (2009). *Identifying and selecting evidence-based interventions revised guidance document for the strategic prevention framework state incentive grant program*. HHS Pub. No. (SMA) 09–4205. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.

Congressional Research Service. (2007). *Federal drug control: Background, legislation, and issues*. Washington, DC: Congressional Research Service Reports.

Crane, B. D., Rivolo, A. R., & Comfort, G. C. (1997). *An empirical examination of counterdrug interdiction program effectiveness*. Alexandria, VA: Institute for Defense Analysis.

Inter-American Drug Abuse Control Commission—CICAD. (2009). *How to develop a national drug control policy: A guide for policymakers, practitioners, and stakeholders*. Washington, DC: Organization of American States.

Kerlikowske, G. (2011). *Addiction is a brain disease and not a moral failing*. Retrieved from: https://ncadd.org/in-the-news/365-addiction-is-a-disease-not-a-moral-failure-kerlikowske

MacCoun, R. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology, 49,* 259–287.

Madras, B. (2010). Office of national drug control policy: A scientist in drug policy in Washington DC. *Annals of the New York Academy of Sciences, 1187,* 370–402.

Manski, C. F., Pepper, J. V., & Petrie, C. V. (Eds.). (2001). *Informing America's policy on illegal drugs: What we don't know keeps hurting us*. Washington, DC: National Academy Press.

Manski, C. F., Pepper, J. V., & Thomas, Y. (Eds.). (1999). *Assessment of two cost-effectiveness studies on cocaine control policy*. Washington, DC: National Academy Press.

Meier, K. (1990). The politics of drug abuse: Laws, implementation, and consequences. *The Western Political Quarterly, 45*(1), 41–69.

Murray, D. (2014). *Hard to study: The difficulty in measuring marijuana's value*. Washington, DC: Hudson Institute. Available at: http://www.hudson.org/research/10604-hard-to-study-the-difficulty-in-measuring-marijuana-s-value

Office of National Drug Control Policy. (2010). *The national drug control strategy, 2010*. Washington, DC: The White House, Office of National Drug Control Policy.

National Drug Intelligence Center. (2011). *The economic impact of illicit drug use on American Society*. Washington, DC: U.S. Department of Justice.

Pew Research Center. (2014). *Views of marijuana—Legalization, decriminalization, concerns*. Washington, DC: Pew Research Center, America's New Drug Policy Landscape. Available at: http://www.people-press.org/2014/04/02/section-2-views-of-marijuana-legalization-decriminalization-concerns/

Reuter, P. (2001). Why does research have so little impact on American drug policy? *Addiction, 96,* 373–376.

Reuter, P. (2013). Why has U.S. drug policy changed so little over 30 years? *Crime and Justice, 42*(1).

Reuter, P., Crawford, G., & Cave, J. (1988). *Sealing the borders: The effects of increased military participation in drug interdiction*. Santa Monica, CA: The RAND Corporation, National Defense Research Institute.

ROPER Center for Public Opinion Research. Retrieved from: http://www.ropercenter.uconn.edu/public-perspective/ppscan/84/84014.pdf

Rydell, C., & Everingham, S. (1994). *Controlling cocaine: Supply versus demand programs*. Santa Monica, CA: RAND, Drug Policy Research Center. ISBN 0-8330-1552-4.

Simeone, R., Carnevale, J., & Millar, A. (2005). A systems approach to performance-based management: The national drug control strategy. *Public Administration Review, 65*(2), 191–202.

Substance Abuse and Mental Health Services Administration. (2014). *Results from the 2013 national survey on drug use and health: Summary of national findings*. NSDUH Series H-48, HHS Publication No. (SMA) 14–4863. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Tieberghien, J. (2012). Understanding the research-policy interface: An analysis of the Belgian debate on cannabis policy between 1996 and 2003. In *International Society for the Study of Drug Policy* (ISSDP-2012). Ghent University, Department of Penal Law and Criminology.

Economic Evaluation of Substance Abuse and Prevention Programs

19

Willie H. Oglesby and Lauren Birmingham

19.1 Introduction

In public health, economic evaluation is the assessment of the economic output of a health intervention compared to the economic inputs required to produce them. Not only do economic evaluations measure *financial* inputs and outputs, which are easier to quantify because they are often actual costs or savings that can be documented, they also measure broader monetary impacts that can be more difficult to quantify, such as quality of life and the monetary value associated with the length of life. The use of various methods available to examine these inputs and outputs is called economic evaluation.

This chapter briefly reviews the main methods of economic evaluation of health interventions focusing on how they can be applied to substance abuse and prevention programs. These include cost-of-illness analysis, cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis.

W.H. Oglesby (✉)
College of Population Health, Thomas Jefferson University, Philadelphia, PA, USA
e-mail: Billy.Oglesby@Jefferson.edu

L. Birmingham
Department of Health Policy and Management, College of Public Health, Kent State University, Kent, OH, USA
e-mail: lbirming@kent.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_19

19.2 Types of Costs

In any economic evaluation, there are generally three types of costs that are examined: direct costs, indirect costs, and intangible costs.

Direct costs are those that can be completely attributable to the disease, condition, or intervention being measured and can be easily quantified. For substance abuse prevention programs, direct costs routinely include salary and fringe benefits costs associated with the exact amount of staff time devoted to the prevention program, consumable supplies used by the program, transportation costs for staff to deliver the program, or other costs that would be directly incurred by the program. For treatment programs, they might include the salary and fringe benefits for the time spend treating clients, costs of medications, and other expenses directly incurred by a treatment program.

Indirect costs are those that cannot be completely attributable to the disease, condition, or intervention being measured, but can be quantified. For substance abuse prevention and treatment programs, this could include costs of administrative and managerial staff (assuming these people are not directly involved in the prevention or treatment program), costs associated with the organization's facilities (such as rent, utilities, and telephone), annual audit expenses for the organization, or other costs that are not directly incurred by the program but are still costs that are incurred by an organization providing prevention and treatment programs.

Because indirect costs cannot be completely attributable to the disease, condition, or intervention being measured, they must be allocated using an appropriate standard. An organization's facilities costs might be allocated based on square footage, where the total costs are divided by the total square footage of the organization and the multiplied by square footage occupied by program staff. Similarly, total telephone expenses can be divided by the number of telephones in the organization and then multiplied by the number of telephones used by program staff. Administrative and managerial staff costs can be allocated based upon the amount of time spent overseeing the particular program or managing program staff.

Indirect costs can also include costs borne by participants. The cost of participant transportation to and from the program should be considered, as well as any lost time spent at work, or doing other productive activities, such as child care. Furthermore, in some treatment programs, family members or others may provide support as part of the program. Their time should also be valued and incorporated into the cost analysis. Valuing indirect costs requires the evaluator to have a strong understanding of what participants would do with their time if they were not participating in the program. This essentially describes the *opportunity cost*, or the value of how a person would have spent their time, had they not been participating in the program.

The concept of an opportunity cost is central to economic theory and highlights the concepts of scarcity and choices. The concept is very simple: the cost of something is what you give up to get it (Mankiw 2012). Consider a simplified example—if someone decides to attend a residential drug treatment program, this means they cannot work during this time. Thus, the opportunity cost of attending a residential drug treatment program is the money the individual would have earned if they would have worked instead of going for residential treatment. This concept is important to understanding how to value the indirect costs of drug abuse programs—because participants give up their time to attend treatment and their time comes at a cost that should be recognized.

Intangible costs are the most difficult to value and are often omitted from analyses. When they are omitted, it should be explicitly stated, so readers can be aware of the omission. The purpose of reporting intangible cost is to estimate the cost of pain and suffering, or emotional pain associated with a health condition. Intangible costs can be valued through more holistic measurement systems such as the *willingness-to-pay* (*WTP*). Other methodologies exist, such as the standard gamble, which are described elsewhere (Gafni 1994).

The WTP method comes from welfare economics and is most frequently used when a market price is not available. WTP essentially asks participants how much money they would be willing to pay in order to be completely disease-free or how much they would pay to reduce symptoms. This helps researchers estimate the total cost of living with a disease—including the intangible costs. For a more in-depth discussion on the WTP method, and other related methods, see Segel (2001).

Direct costs	Indirect costs	Intangible costs
Program operations (e.g., program employees, materials, equipment, etc.) Fixed costs (e.g., rent for space, etc.) Overhead costs (e.g., utilities, secretarial services, etc.)	Patient transportation to/from program Lost productivity due to program participation Caretakers time	Cost associated with decreased quality of life associated with disease/condition (pain, suffering, etc.) Stigmatization associated with condition

Table adapted by writing from French (1995)

19.3 Perspective

The context in which costs are analyzed (and eventually benefits) is referred to as *perspective*. In economic evaluation, perspective is the vantage point taken when analyzing the economic costs and benefits of a disease, condition, or intervention. These vantage points usually differ across the various stakeholders who have an interest in the disease, condition, or intervention under study.

For example, a state legislator who is interested in allocating funding for a new state-wide heroin overdose prevention initiative would need to consider the costs associated with the prevention program, cost reductions to state programs like Medicaid resulting from the prevention program, the benefits that those dollars would have generated if they were spent on something else (i.e., the opportunity cost), or other costs and benefits that impact the government. This perspective is generally known as the *government perspective* since the focus is on the costs and benefits to the government. It is a common perspective in substance abuse programs, since the government is a large payer of these services. However, it is not the only payer or perspective one can take in economic evaluation of substance abuse programs.

Consider the example of an insurance executive who is trying to assess the economic impact of a new substance abuse treatment program that her company would like to offer during the next plan year. In this perspective, the insurance executive might only be interested in the direct costs and benefits to the insurance company. It would not include savings to the Medicaid program (unless the insurance company was a Medicaid managed care company, of course), or any other cost or benefit to the government for that matter. In this perspective, he/she will likely only be interested in the costs associated with the additional claims to be paid to treatment providers for this new service, the slight increase in cost associated with processing more claims, and negotiating rates with providers, cost savings to the insurance company resulting from reduced hospitalizations, or other costs and benefits that directly impact the insurance company. This is generally known as the *insurance perspective* and is another common perspective in substance abuse programs, since more insurance companies are paying for substance abuse services as a result of the Affordable Care Act and recent mental health parity laws.

Finally, consider the perspective of society as a whole. Assume you are a health policy analyst who is interested in examining the costs and benefits of a national school-based substance abuse prevention program. Using this perspective, we would be interested in all costs associated with providing the intervention including the costs for participating, such as transportation costs for parents, the cost associated with parents' time away from work (to transport their kids or if they are involved in the intervention), or other costs of implementing and participating in the intervention. This is the most comprehensive and inclusive perspective taken in economic evaluation and is generally known as the *societal perspective*. It is also one of the most difficult to completely assess because of the very long list of indirect costs that it encompasses.

Ultimately, the perspective applied to an economic evaluation should depend upon the research questions being asked and how the results will be used. Research questions that drive decision-making at a governmental level will usually apply a government perspective to the analysis. Financial decisions at an insurance company will likely be limited to the costs and benefits incurred by the company. Larger policy-level analyses will take a more comprehensive societal perspective.

Although different perspectives can be taken in various economic evaluations, health policy experts recommend taking a societal perspective for economic evaluations designed to address broad problems. For a more detailed explanation and justification of this, as well as other recommendations, see Cost-Effectiveness in Health and Medicine (Gold et al. 1996). Defining the perspective and calculating, costs are a prerequisite to any economic evaluation. Having these tools allows the reader to now learn about the four types of economic evaluations.

19.4 Cost-of-Illness Analysis

Cost-of-Illness (*COI*) *analysis* seeks to measure the total financial burden of a disease or condition. This can be expressed in terms of the financial burden on the patient, the patient's family, state budgets, insurance companies, society as a whole, or some other perspective. At times, researchers will take different perspectives because the total financial burden of a disease or condition is different for that particular vantage point, or payer.

In a cost-of-illness analysis, the focus is just on identifying the costs associated with having a particular disease or condition and does not examine any cost savings or other benefits resulting from a health intervention that can be administered to treat or cure it. Cost-of-illness analyses are conducted specifically to better understand the total financial burden of having the disease or condition, such as addiction, which can be used for internal budget planning, larger resource allocation decisions, or to use in more advanced economic analyses (discussed later).

19.4.1 COI Methods

There are two primary methods used in COI estimations—prevalence and incidence approaches. Of the two, the prevalence method is most common. This approach estimates the total cost of a disease in a given year, whereas the incidence approach estimates the cost of the number of new cases diagnosed in a given year. The challenge lies in determining which factors to consider in this cost. Perspective is thus critical. From a societal perspective, for instance, all direct and indirect are considered. For example, in the case of substance abuse, these might include direct medical care expenditures (including travel costs to and from the clinic, etc.) as well as indirect costs such as value of lost productivity for people who are ill or who die prematurely due to their substance abuse. The cost associated with this lost productivity is typically determined by the human capital approach, which assigns value quantified in terms of forgone earnings, standardized for age and sex.

19.4.2 Limitations of Cost-of-Illness Analyses

While COI studies are intended to identify and measure the costs of a health problem as a necessary "first step" in the prioritization of health policies and interventions, they have been criticized as a measure of condition severity, based on their inconsistencies in measurement and inability to fully enable choice considerations (Drummond 1992; Wiseman and Mooney 1998). Variability in measurement of costs has been the major issue. This variability primarily arises based on three considerations: (1) choices made in terms of the actual costs considered in a given analysis (partially determined by the perspective chosen for a given analysis), (2) inconsistencies across data available to estimate costs due to variations in the methods employed across studies, and (3) the inability to disaggregate all relevant costs (Clabaugh and Ward 2008).[1] While not intended as an exhaustive list, there are other inherent limitations as well. According to Shiell et al. (1987), for example, another important barrier to the application of COI analysis is that it cannot address the most pertinent question facing decision-makers, which is one of scales—the degree to which programs should be expanded or contracted. These answers necessarily require some ability to compare expected change in benefits associated with costs, which COI analysis alone cannot inform (Byford et al. 2000; Currie et al. 2000; Shiell et al. 1987). Finally, it is simply challenging to accurately establish a monetary value for all potential factors involved.[2] Included in this

[1]Following their systematic review of methodologies used in cost-of-illness studies in the United States, Clabaugh and Ward (2008) urged caution when interpreting or applying results, until such time that accepted standards for the execution COI studies were adopted.

[2]While a detailed discussion of the human capital approach is beyond the scope of this chapter, there are a number of issues with the estimates derived. Most

challenge is the tendency for COI analyses to ignore the concept of local rationality, where the more proximate policy and practice context may be at odds with the broader societal perspective in assigning value (Brouwer et al. 2006).

All of these limitations are evident in COI estimates of substance abuse. Specifically, while intending to assess the array of substance abuse-related costs in areas such as health care, productivity, and criminal justice, COI analyses consistently result in estimates biased toward harms (Murphy et al. 2015). This is in large part because the costs associated with negative outcomes are relatively easier to identify measure (Kleiman 1999). Thus, the resulting estimates fail to fully take into account for all associated societal costs, such as the violence and family dysfunction that typically accompanies substance abuse, which while difficult to measure may actually be significant in any true valuation (Kleiman 1999). Even if identified, the magnitude of costs associated with substance abuse is still often difficult to quantify, in part because substance abuse may be only one risk factor associated with the harms identified (Murphy et al. 2015).

19.4.3 Case: Exploring the Economic Costs of Drug Abuse

While scientists have sought to articulate the economic costs of substance abuse, how costs are calculated has significant implications for study results and application. A review by Cartwright (2008) explored the complexity of using cost-of-illness methodology from multiple perspectives. Specifically, he examined the volatility of using health plan data, as well as limitations in using premium and parity cost estimates to inform policy decisions. Calculations from a societal perspective are also challenging, particularly due to the sheer number of component costs associated with all health and non-health outcomes that must be identified and measured when considering all everyone affected by drug use.[3] Finally, costs estimates derived from more geographically limited sources were considered. Once again, key challenges were delineated, particularly related to concerns about the applicability and/or generalizability of data when derived from novel programs in terms of target population(s), location, and/or service provision.[4] The interests of different stakeholders associated with more localized studies also play important roles in cost calculations. Cartwright concludes by noting that estimates of the value of drug abuse insurance coverage across society remain fundamental. However, improved costing approaches and methods need to be developed to better measure individual and societal costs associated with drug abuse, including estimates that integrate the costs of drug abuse treatment services with ancillary services to provide a more accurate depiction of individual program and system costs.

19.4.4 Conclusion: Cost-of-Illness

Cost-of-illness analyses are focused just on the costs associated with the illness, with the goal of measuring the total financial burden of that disease or condition. The scope of our costs is adjusted by the perspective taken, whether the focus is just on the patient or the patient and their family, the impact on state budgets and/or insurance companies, the cost to society as a whole, or some other perspective. The broader the perspective, the larger the cost-of-illness will

(Footnote 2 continued)
notably, because it relies on existing earnings patterns, greater weight is given to working-aged men compared to other demographic groups.

[3]Examples of costs include nonmonetary costs, such as emotional distress and productivity losses associated with the illness, premature death, and crime associated with drug use. Equally important are the treatment costs associated with the many medical complications associated with abuse, estimated to amount to US$8.4 billion (Cartwright 2008).

[4]For example, estimates for counseling depend on the nature of the patient, as well as the type of treatment program under consideration. In one residential program, cost of an hour of an hour of counseling was about two-thirds less for long-stay patients compared to short-stay patients (Alemi et al. 2002).

be because of the additional costs it will include. However, the broader the perspective, the more difficult it is to quantify the costs and the more challenging it will be to defend because of the assumptions the evaluator will need make. A narrower perspective will be easier to quantify and justify, but it will not include additional, often indirect and intangible, costs that are known to contribute to the total financial burden of the illness. Ultimately, the economic evaluator will adjust their perspective and judgments based on generally accepted techniques found in the scientific literature and his or her experience.

19.5 Cost-Effectiveness Analysis

As new substance abuse prevention and treatment strategies emerge, it is important to assess its cost-effectiveness as well as the overall program effectiveness. After the new health intervention has been determined effective at impacting a particular set of outcomes, the cost-effectiveness of that intervention against current best practices or perhaps other intervention options should be examined. To do this, the *cost-effectiveness analysis* approach is utilized. In a cost-effectiveness analysis, or CEA, the costs and health outcomes of a particular health intervention are compared with the costs and health outcomes of some other comparator. The other comparator is usually the "current standard of care," but it can also be other similar health interventions, as well.

19.5.1 Choosing the Health Outcomes and the Comparator

The outcomes of health interventions for substance abuse can range from more immediate changes, such as denormalization of substance use, increased resiliency, treatment readiness, and therapeutic engagement to longer term changes, such as prevented or delayed onset of substance use or sobriety maintenance. Distal outcomes such as increased length and quality of life can also be used, but when they are, the evaluation is generally called a cost-utility analysis (described later). In a CEA, the outcomes of health interventions to be compared must be the same. It could be any of the health outcomes listed above, or others, as long as they are directly caused by and are the same across the health interventions to be examined.

As an example, consider two modalities for individuals undergoing treatment for drug addiction. The first intervention includes counseling and behavioral therapies and the second intervention includes the same counseling and behavioral therapies, plus a prescription medication for drug addiction. The outcome of interest is the number of overdoses within 12 months of beginning treatment. In this case, both groups (counseling + behavioral therapies group, and the counseling + behavioral therapy + drug group) would be tracked for 12 months, and the number of times they overdosed would be measured.

19.5.2 Choosing the Comparators

In CEA, the intervention being assessed must be compared to other alternatives. The alternatives can include the "current standard of care," other intervention strategies that have the same identified outcomes, and no intervention at all. Since most economic analysis is about making decisions among several choices, the comparators must be realistic alternatives to the intervention being assessed.

19.5.3 Incremental Cost-Effectiveness Ratio (ICER)

As noted, in a CEA, researchers are frequently comparing the cost-effectiveness of one intervention to others. The "other" interventions usually include the current standard or care, other similar interventions, and no intervention at all. As a result, a cost-effective analysis usually ends

up comparing an intervention to several others with the goal of identifying how much "better" the intervention is compared to the other alternatives. To do this objectively, the incremental cost-effectiveness ratio (ICER) is used. The incremental cost-effectiveness ratio calculates the difference in costs and outcomes between the intervention and a comparator and calculating a ratio where costs are in the numerator and outcomes are in the denominator. In its most basic form, this is illustrated using the formula below:

$$\frac{(\text{Total Costs}_{\text{intervention}} - \text{Total Costs}_{\text{comparator1}})}{(\text{Outcome}_{\text{intervention}} - \text{Outcome}_{\text{comparator1}})}.$$

An ICER should be calculated for each comparator and the results should be presented in tabular form with the lowest ratio at the top. The intervention with the lowest ratio has the lowest cost per outcome gained and thus, is the most cost effective compared to the other alternatives.

19.5.4 Limitations of Cost-Effectiveness Analyses

In comparing costs to physical benefits, cost-effectiveness analyses avoid many of the measurement challenges associated with attempting to directly assign monetary values to life and health. However, conclusions drawn still share many disagreeable features with other economic evaluations, essentially pushing cost considerations onto the decision-maker. Estimations are also still subject to the influence of inconsistencies in income estimates and data variability. Other considerations are also still operative. In particular, the estimates derived from this methodology, in and of themselves, are poor measures of efficiency, as costs and benefits are measured in different units and therefore do not lend themselves readily to the creation of a measure of net benefits. More importantly, cost-effectiveness analysis is not an ideal methodology to rank options across dissimilar programs, such as a comparison of the benefits of a methadone program to those of a high school educational intervention, limiting its usefulness as a single indicator for priority setting.[5]

Finally, cost-effectiveness analysis has been criticized as being an oversimplification of what are often highly complex processes, primarily through what many consider a rather arbitrary threshold for determining cost-effectiveness (Diamond and Kaul 2009). Convention has typically set $50,000 per quality-adjusted life year as the socially acceptable threshold for establishing cost-effectiveness. The assumptions behind this threshold, as well as other guidelines for cost-effectiveness analysis, have been panned as having little basis in contemporary economics (Brouwer et al. 2006; Diamond and Kaul 2009). To illustrate, critics often point to the fact that the established threshold for effectiveness has remained largely static over decades of work, despite considerable change in the economic reality in which decision-makers must operate.

19.5.5 Case: Cost-Effectiveness of Treatment for Alcoholism

Policy discussions in the U.S. have long focused on the need and/or value of having health insurance benefits covering alcoholism treatment. In order to begin to answer this question, Holder et al. (1991) sought to review and assess the cost-effectiveness of different treatment modalities (e.g., Alcoholics Anonymous, psychotropic medications, residential therapy, etc.) for alcoholism. In this ambitious endeavor, they note up front some of the key challenges, including lack of consensus among researchers or clinicians with regard to the independent and interactive effects of different treatment modalities or a commonly accepted standard of effect,[6] and a

[5]The same can be said of any effort to prioritize based on a single criterion (Baltussen and Niessen 2006). The reality is that the situations are often far more complex. Therefore, priority setting efforts must take into account multiple criteria simultaneously to be effective.

[6]While alcoholism treatment does not have a commonly accepted standard of effect, in many situations, the intended treatment goal is abstinence. However, when

lack of standardization in measurement of treatment costs. Using carefully selected rules of evidence, the authors still sought to integrate cost and effectiveness studies published in the literature. In doing so, they noted some of the key issues associated with cost estimates for alcoholism treatment, including significant variations by setting (e.g., inpatient, residential, etc.)[7] and treatment modality. Treatment modalities were also classified, based on scoring procedures, into five qualitative categories differentiating evidence of effect, controlling for intensity of treatment modality. Their analyses provided evidence that high-cost treatments were not necessarily associated with better outcomes. In fact, none of the modalities with "good evidence of effectiveness" reviewed placed in the "medium-high" or "high" cost categories. On the other hand, at least one of the modalities in the "high" cost category was classified as having "no evidence of effectiveness."[8] While not definitive, available evidence from this first approximation encourages utilization of lower cost modalities as a means to more effectively treat alcoholism until more evidence is available supporting pricier inpatient treatment options.

19.5.6 Conclusion: Cost-Effectiveness Analysis

In cost-effectiveness analysis researchers compare one intervention to other alternatives by examining the differences in costs and outcomes. In CEA, the outcomes of the interventions must be the same and each comparator must be realistic alternatives to the intervention being assessed. The "additional cost per outcome gained" is expressed as the incremental cost-effectiveness ratio where the lowest ratio is the most desirable among the alternatives. Additional years of life and improved quality of life can be used as outcomes, but when they are, the analysis is generally considered a cost-utility analysis (described next).

(Footnote 6 continued)
measuring abstinence, the necessary length of time a subject abstains from using alcohol to qualify is debated.

[7]Costs differences across setting were due, in part, to the different staffing requirements associated with each.

[8]It should be noted that insufficient evidence is not equivalent to "not effective" in this case. More evidence is needed before the higher expenditures could be sufficiently justified.

19.6 Cost-Utility Analysis

One of the reasons economic evaluation is necessary is to help decision-makers decide which programs or interventions they will support, given limited resources. Cost-utility analysis (CUA) is thought to be one of the better means of analysis for allocating resources (Robinson 1993a; Vanhook 2007). The primary benefit of CUA over CEA is that cost-utility analyses allow people to compare multiple interventions at once —even when the health outcomes are measured differently. This is helpful to decision-makers, as they often have a limited budget and can only select a few programs to fund.

19.6.1 Cost-Utility Analysis and Cost-Effectiveness Analysis: Similarities and Differences

Cost-utility analysis and cost-effectiveness analysis are similar in that costs are measured in the same way for both forms of analysis. The differences arise in how health benefits are measured. In cost-effectiveness analysis benefits are measured using clinically significant metrics—such as drug overdose incidents averted. In cost-utility analysis, benefits are measured using the standardized metric called QALYs—or Quality-Adjusted Life Years. While health outcomes need to be measured in order to conduct either type of analysis, CUA goes one step further by standardizing the measure of benefit into QALYs.

19.6.2 QALYs

Quality-adjusted life years, or QALYs (pronounced "qwaleys"), are a critical concept in

understanding cost-utility analysis. A QALY essentially represents how many additional years of life, adjusting for quality of life, are gained by participating in a program or intervention. This concept takes into account that treatments have side effects, and while mortality may have been prevented because of the treatment, morbidity may have been introduced into the life of the participant because of the treatment. Another similar concept is used in the literature called disability adjusted life years, or DALYs. The DALY measures the length of life as adjusted by the effect of disability. A QALY is a mathematical combination of years of life saved due to the treatment or intervention that is adjusted by a factor that takes into account the quality of life a person endures after treatment. One year of perfect health is measured as 1 QALY, and death has a value of 0 QALYs. The values in between 0 and 1 represent less than perfect health.

Measuring quality of life can be a challenging task and requires the use of instruments that measure both objective and subjective attributes. Measuring quality of life has become increasingly important, as the focus of many public health efforts in the United States has moved from infectious diseases to chronic diseases. In the case of chronic diseases, measuring mortality does not adequately evaluate how well chronic diseases are treated. Certainly, reducing death associated with chronic disease is good—but prolonging a life that is full of pain and suffering is not entirely desirable. For this reason, it is necessary to measure quality of life as well. Many tools exist that seek to measure what is known as *Health-Related Quality of Life* (*HRQL*). HRQL seeks to paint a comprehensive picture of overall health and well-being. HRQL tools measure multiple facets of life including physical and mental health, mobility and functional status, and other related metrics.

A number of tools have been developed to measure quality of life including the Medical Outcomes Study Short Forms (SF-12 and SF-36) (www.cdc.gov/hrqol/concept.htm), EQ-5D (www.medicine.ox.ac.uk/bandolier/painres/download/whatis/QALY.pdf), and the Health Utilities Index (HUI) (www.hqlo.com/content/1/1/54). These questionnaires ask questions about how participants rate their general health, how their health impacts their daily lives, and to what degree their health prevents them from participating in regular activities. The overarching purpose of the HRQL surveys is to determine how many years of unhealthy life are equal to 1 year of perfect health. This establishes a factor (less than 1) that can be used to adjust the number of life years saved by an intervention. Once the results have been tabulated and a score is determined, the number of years of life can be adjusted by the factor.[9]

19.6.3 Comparing Multiple Interventions

One of the primary advantages associated with cost-utility analysis is that it provides a standard way to compare multiple interventions or programs. When QALYs and costs are used together, the amount of money needed to "buy" one QALY can be calculated. For example, Intervention A provides 0.8 QALYs for 5 years and Intervention B provides 0.6 QALYs for 5 years. At the end of 5 years, Intervention A generates four QALYs and Intervention B generates three QALYs. If it also known that Intervention A costs $50,000 and Intervention B costs $10,000, then the cost per QALY for Intervention A is $12,500 and the cost per QALY for Intervention B is $3333. Calculating the cost per QALY can be helpful in comparing multiple interventions because it provides a standardized way to look at health outcomes and costs.

	Intervention A	Intervention B
QALY	0.8	0.6
QALYs in 5 years	(0.8*5) = 4 QALYs	(0.6*5) = 3 QALYs

(continued)

[9]For more information on how health-related quality of life is measured in public health practice, see Hennessy et al. (1994).

	Intervention A	Intervention B
Cost of intervention	$50,000	$100,000
Cost per QALY	$12,500 = ($50,000/4 QALYs)	$3333 = ($10,000/3 QALYs)

One challenge decision-makers deal with is deciding what the threshold is for determining which interventions provide enough value to fund, and which do not. Said another way, decision-makers need to decide the minimum cost per QALY they will accept, or what the amount is that they are willing to pay for one QALY. Is $15,000 per QALY acceptable? What about $50,000 per QALY? $150,000 per QALY? Other factors may be taken into account—such as the severity or prevalence of disease, or the population whom the disease effects (McCabe 2009). A $50,000 per QALY threshold has been established by the literature, although this figure is not set in stone for all disease types or populations (Neumann et al. 2014).

19.6.4 Limitations of Cost-Utility Analysis

While cost-utility analysis provides a useful tool for decision-makers, it is not a perfect tool for evaluation, and the following potential disadvantages should be considered in interpreting cost-utility analyses:

1. The analysis is only as strong as the sum of its parts—thus, if the tool used to assess health-related quality of life does not accurately measure HRQL for the patient population, then the results will have decreased validity. When conducting a HRQL it is best to use a validated tool, such as the SF-series, EQ-5D, or HUI.
2. Cost-utility analyses typically do not take into account the impact on quality of life for those who take care of the patient in the analysis. This can be a shortcoming on this form of analysis and should be noted.
3. There is limited generalizability beyond the country where the analysis is conducted due to differing prices for medical services and exchange rates.
4. Limited usability for chronic conditions:
   (a) The quality of life is often more of interest for chronic conditions, rather than survival, since many people live with chronic conditions over the course of a normal lifetime. This makes measuring the quality of life accurately imperative. One other shortcoming is that HRQL measurements assume the health state is the same overtime. This is often not the case with chronic diseases like multiple sclerosis (MS), which is a progressive condition that worsens overtime, thus potentially decreasing quality of life overtime. This can weaken estimates for quality of life.
5. Limited usability for preventative health screenings
   (a) Measuring the benefits of preventative health screenings is challenging because the impact of the health screening may not be observed for many years—or perhaps never at all. While this may be a difficult task, QALYs have been calculated for preventative health screenings, including HIV screenings (Dowdy et al. 2011).

19.6.5 Conclusion: Cost-Utility Analysis

Cost-utility analysis (CUA) is a useful tool for decision-makers with limited budgets who need to get the most "bang for the buck." Cost-utility analyses allow decision-makers to compare multiple interventions whose focus may be varied across clinical or practical areas of public health.

The CEA Registry, a registry maintained by Tufts University of cost-utility analyses, provides a library of cost-utility analyses examining a variety of interventions. This allows researchers and practitioners alike to examine the effectiveness of interventions in standardized units.

CUA is similar to CEA, except benefits are measured by QALYs rather than in clinical terms. A QALY is the number of years of life saved by an intervention, adjusted by a factor representing potential decreases in quality of life painting a more accurate representation of the benefits of treatment. While CUA is a very useful evaluation tool, it does not come without limitations. CUA relies on valid assessments of quality of life, so QALYs can be calculated and may have somewhat limited usability in assessing the value of treatments for chronic conditions and preventative health screenings.

19.7 Cost-Benefit Analysis

Cost-benefit analysis is one of the most common forms of economic analysis because the results are the most easily understood by lay persons (even though it is one of the more difficult methods to master!). The process essentially consists of measuring both the financial costs and financial benefits associated with an intervention, and then calculating whether there is a net gain or net loss in welfare. There are many challenges and nuances associated with this form of analysis that will be discussed, including the different types of benefits and how to measure them, how to measure costs in dollars, and how to value them over time.

Cost-benefit analysis is critical in a time when resources dedicated to healthcare programs are very limited. Decision-makers need to be given "actionable information," or information that can be used effectively in the decision-making process. From the perspective of the program manager, being able to quantitatively document a program's success in an economic evaluation is critical to receiving funding. Showing that a specific program can provide more net benefit to society than other programs will likely improve the odds of financial support when decision-makers start to allocate money for programs. While cost-benefit analysis is not the end-all-be-all in decision-making, it certainly is a tool that can be used in assessment. However, cost-benefit analysis does not have a way to evaluate the ethics or morality of treatment options (Plotnick 1994)—thus these evaluations need to be conducted using different methodologies.

19.7.1 Measuring Costs

The first step in conducting a cost-benefit analysis is to examine the costs associated with the intervention. As described previously, costs can be broken into three types—direct, indirect, and intangible costs. *Direct costs*, or the accounting cost, refer to the costs of providing the intervention (employees, supplies, equipment, etc.). *Indirect costs*, or the economic costs, refer to productivity losses due to participation in the program. *Intangible costs* refer to costs associated with pain and suffering, or decreased quality of life. Refer back to the cost-of-illness section for more detailed discussion of these costs.

19.7.2 Measuring Benefits

The next step is measuring the financial benefits associated with the intervention. The benefits will also be direct, indirect, and intangible—but must be expressed in monetary terms. As previously discussed, direct benefits are more easily measured than indirect benefits. Direct benefits can be observed by examining pre-/post-program participation medical expenditure data for participants, reduced expenditure on drugs by program participants, reduced police and court costs associated with illicit drug possession, less staff time needed to address student behavioral and academic issues, lower costs in the mental and behavioral health care system, or other similar costs. These benefits could also be longer term such as higher earnings in adulthood resulting from higher academic achievement and avoiding

addiction, increase in the quality and length of life, and other longer term benefits of substance abuse prevention programs. Direct benefits are easily observed and can be quantified with relative ease.

In contrast, indirect benefits include increased productivity by program participants (e.g., perhaps the successful participant is able to spend more days at work per month), improved quality of life, and increased public safety. Indirect benefits are not as easily observed, and can be more difficult to quantify. Another way to think about indirect benefits is to think of them as positive spillover effects due to the intervention that, while beneficial to the participant or society, are not the purpose of the intervention. For example, an indirect benefit of participating in a drug treatment program may be increased worker productivity. While the goal of the drug treatment program is not to create a more productive workforce, it can be a side effect of successful program participation. These benefits are more difficult to measure; however, they provide important insight into the breadth and depth of benefits that can be accrued as a result of an intervention. Common benefits of substance abuse prevention programs can include the following:

Direct benefits	Indirect benefits
Lower healthcare costs due to reduced or abstained drug usage Reduced legal costs Decreased absenteeism/increased presenteeism by both participating and family members or friends providing assistance	Improved quality of life Increase in productivity Increased public safety (reduced criminal activity)

Table adapted by writing from French (1995)

19.7.3 Calculating Cost-Benefit Ratios

After costs and benefits are calculated, they are compared using the cost-benefit ratio. The cost-benefit ratio essentially illustrates how much "bang for your buck" a certain program produces (which is the primary reason why it is favored by policymakers). The ratio presents the dollar value of benefit gained per dollar spent on the program.

One advantage of calculating cost-benefit ratios is that they are very easy to compare across programs with different outcomes, since the common denominator is in monetary terms. However, ratios can be somewhat misleading as the time it takes to realize the stated benefit is not necessarily common across all cost-benefit ratios. One program may produce $2 of benefit for every $1 spent and generates financial benefits immediately, while another produces $5 of benefit for every dollar spent but takes 10 years to achieve any amount of benefit. Thus, while the cost-benefit ratio is a useful way to evaluate programs—it is not an end-all, be-all form of analysis. To help understand the impact of time, the present value of money can be calculated to help make informed decisions that include time as a consideration.

19.7.4 Present Value Analysis

Cost-benefit analysis requires measuring the value, in dollars, of costs and benefits over time. This frequently includes assessing the value of benefits received in the future and assessing costs that have been incurred in the past. In order to compare these costs fairly, they must be adjusted, with the final product called present value analysis.

Adjusting future values to present terms is called discounting. The idea of discounting relies on the assumption that a dollar received today is worth more than a dollar received a year from now. This is the case because a dollar received today could theoretically be invested and earn interest over the course of a year, and be worth more than a dollar at the end of one year—whereas a dollar received a year from now is still worth only a dollar. Essentially, the concept of discounting reflects the opportunity cost of holding money. Rosen and Gayer (2008) define

the present value concept as "the value today of a given amount of money to be paid or received in the future."

Given the fact that future amounts of money are worth less today, we must discount future amounts of money to determine their present value. Determining the present value is accomplished through a process known as discounting. In order to calculate a present value, a discount rate, r, must be assumed. The discount rate is the assumed rate of interest that could reasonably be obtained in the market.

$$\text{Present value formula: PV} = \frac{\text{FV}}{(1+r)^n}$$

where

FV	future value
PV	present value
t	time period (usually in years)
r	discount rate.

An example best illustrates how this formula works. Pretend the future value of a program is thought to be $50,000 in 3 years. Furthermore, assume a discount rate of $r = 0.05$, or 5%:

$$\text{PV} = \frac{\$50,000}{(1+0.05)^3} = \$41,192.$$

This means that the present value of $50,000 in costs or benefits received 3 years from now is worth only $41,192 in present value terms. Discounting future values is a common and very important practice. Forgoing the discounting process can lead to serious overestimates of the value of both benefits and costs, which would bias the end decisions. Likewise, when historical costs are used, they must be increased to the present terms because when costs are compared over time, they must be in the same period so that resulting analyses do not over- or underestimate their true value. To bring historical costs to their present values, we must consider inflation that occurred over that time period.

Inflation is the general increase in the price of goods and services over time. In the United States, the Bureau of Labor Statistics tracks changes in these prices using the Consumer Price Index (CPI). The CPI is a market basket of goods and services purchased by US consumer household and the costs of those goods and services are tracked over time. The change in those prices serves as a general measure of inflation and is commonly used in economic analysis.[10]

It is important to note that the market basket of goods and services tracked for the CPI includes medical costs. In fact, medical care is one of eight major groups in the CPI. As such, researchers can obtain inflation statistics just for the medical care component of the CPI, which is the preferred method for adjusting medical costs. Medical inflation has generally been higher than overall inflation over the last few decades, so using medical inflation percentages to adjust medical costs will produce better, and more defensible, results. Salaries and costs of other day-to-day living expenses can be adjusted using the overall CPI.

19.7.5 Sensitivity Analysis

A sensitivity analysis is conducted to show how variation in certain assumptions causes the end result to change. Sensitivity analyses show how responsive certain variables or assumptions are to small changes. For example, in a cost-benefit analysis the assumption of the discount rate used in calculating the present value of costs and benefits could be tested in a sensitivity analysis.

Gentilello et al. (2005) present a graphical depiction of a sensitivity analysis for an emergency-department-based alcohol intervention program. As shown in Fig. 19.1, the authors produced a tornado plot—graphically displaying the variations in cost savings associated with their key metrics. As you can see, the values in each category vary to some degree which can have an impact on the ultimate cost-benefit outcome. Presenting the results of a sensitivity analysis helps the reader understand the assumptions behind the model and understand

[10]For more information on the CPI, visit the Bureau of Labor Statistics web site http://www.bls.gov/cpi/.

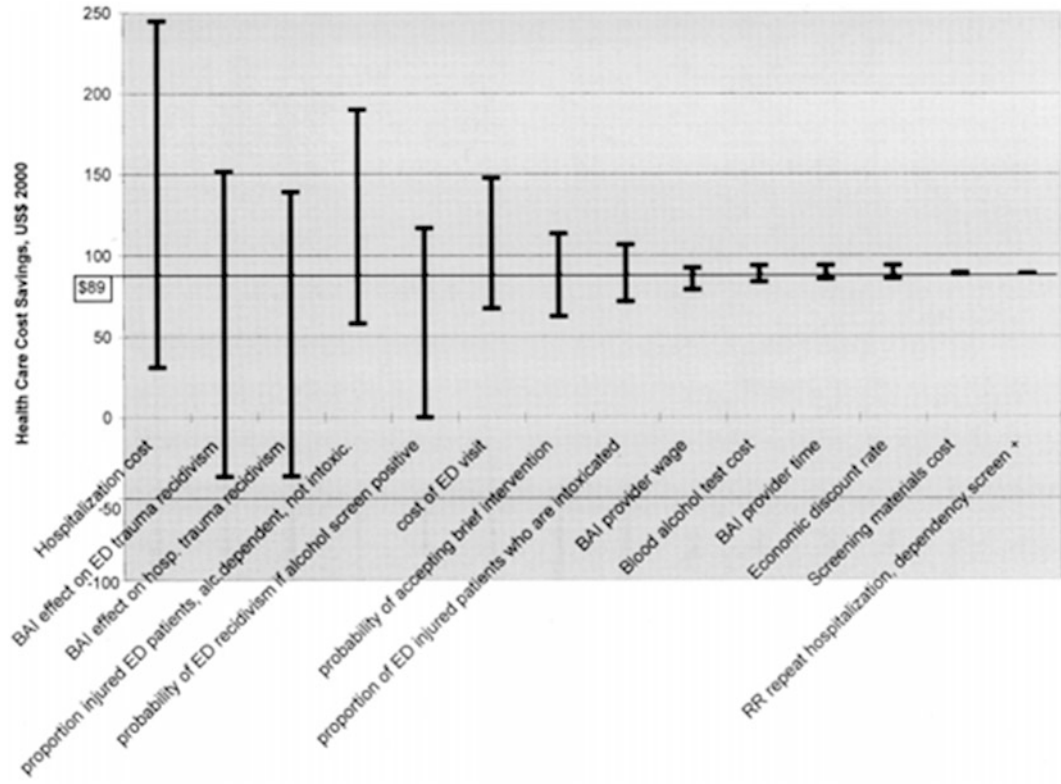

Fig. 19.1 Sensitivity analysis demonstrating potential cost savings associated with variations in the number of primary variables. Screening and brief intervention is associated with cost savings when the bar is above zero. The estimated man cost savings is $89 in US dollars (year 2000). BAI, brief alcohol intervention; ED, emergency department; RR, relative risk. Figure reproduced from Gentilello et al. (2005)

how sensitive the components of the model are to change.

19.7.6 International Considerations

When evaluating literature from outside the United States, the reader should be cautious in applying the findings to programs in the U.S. medical costs differ across countries, as does the value of currency. Andresen and Boyd (2010) present cost-benefit and cost-effectiveness analyses for a supervised injection facility for injection drug users in Canada. However, some of their cost data were available in US dollars and other in Canadian dollars. This analysis provides an example of how costs are translated into one common currency for evaluation—but still taking into account the fact that treating the same disease in two different countries is associated with two different costs—even when currency changes are controlled for in the evaluation.

19.7.7 Limitations of Cost-Benefit Analysis

Cost-benefit analysis has been described as the most theoretically sensible and complete form of economic evaluation, in part due to its ability to place monetary values on both inputs (costs) and outcomes (benefits) associated with an intervention or program (Robinson 1993b). Correspondingly, inherent limitations of this method deal primarily with the method employed to assign value. Early criticisms of cost-benefit analysis

were primarily related to the use of the human capital approach to assign valuations. As previously noted, this method has a number of limitations. This early method, although not entirely replaced, has been largely superseded by more sound estimation methods based on individual's stated preferences. However, there remains a number of practical challenges to assigning dollar values to health and life. One key issue is related to whose benefits are to be counted (Boardman et al. 2011; Trumbull 1990; Whittington and MacRae 1986). Numerous populations (undocumented immigrants, criminals, illicit substance users, etc.) are sometimes problematic, in so much as they often do not have full "standing," such that they are allowed to participate in the decision process by having their preferences counted. Problems of standing also arise in the valuation of life, especially in considerations of future generations. Issues associated with standing limit the utility of cost-benefit analysis to issues where there exists some level of consensus with regard to the value assumptions assigned. Other limitations are philosophical.[11] For example, as previously noted, cost-benefit analysis does not have a way to evaluate the ethics or morality of treatment options, as positive consequences of illegal acts are counted in the same way as consequences associated with socially acceptable ones. Bias also poses challenges, as researchers, while seeking inclusiveness, typically only include consequences that can be identified and assigned value.

19.7.8 Case: Cost-Benefit of SBIRT from an Employer's Perspective

Screening, Brief Intervention, and Referral to Treatment (SBIRT) programs are a comprehensive intervention, developed for both clinical and community-based application, designed to identify, reduce, and prevent substance abuse.[12] In order to better understand the costs and benefits of SBIRT, Quanbeck et al. (2010) conducted a cost-benefit analysis from an employer's perspective, as part of an evaluation of SBIRT services delivered in 20 primary care clinical settings in Wisconsin.[13] This analysis used simulation modeling to assessed two types of productivity losses: absenteeism (estimated based on the daily wage rate of the absent worker, with a multiplier to account for additional costs associated with an employee missing work) and impaired presenteeism (based on empirical estimates of costs related to chronic health conditions associated with alcoholism). Screening and treatment costs were initially fixed at $247 per employee, but were reduced in the final model to $227 after discounting. After discounting and adjustment for estimates of staff turnover, absenteeism costs were reduced by $175 per employee, compared to an $823 reduction per employee in presenteeism costs, resulting in a net present value of $771 per employee after accounting for the fixed screening/treatment costs. The results of this simulation modeling of potential productivity benefits suggest that the advantages accrued by employers due to SBIRT screening would be both positive and substantial.

19.7.9 Conclusions: Cost-Benefit Analysis

In cost-benefit analysis, the costs of an intervention are compared with the financial benefits gained and comparing those costs and benefits across different alternatives. In CBA, the costs and benefits are expressed in monetary terms at the same time period—using present value analysis as appropriate. The interventions do not have to have the same health outcomes because

[11] See Hansson (2007) for a comprehensive review of the philosophical problems associated with cost-benefit analysis.

[12] For additional information, see www.samhsa.gov/sbrit/about.

[13] SBIRT has been previously demonstrated to have cost-benefit from a societal perspective (Fleming et al. 2002). In this analysis, the authors wanted to answer from an employer's perspective: Should employers be willing to pay for SBIRT services?

the denominator is the same: money. It is a favored method of policymakers due to the simplicity of how to interpret the findings. However, calculating indirect benefits is very challenging and often requires advanced techniques.

19.8 Chapter Summary

This chapter has endeavored to provide a brief introduction and overview of the major methods used in economic evaluation for substance abuse prevention programs. These include cost-of-illness analysis, cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis. In cost-of-illness analysis, the goal is to calculate the financial burden of a disease or condition. In cost-effectiveness analysis, the goal is to compare the incremental costs and outcome gains to alternatives that have the same outcomes. Cost-utility analysis is essentially identical to cost-effectiveness analysis except that the outcomes are quality of life or quality-adjusted life years. Finally, in cost-benefit analysis the goal is to identify which among a diverse set of programs achieves the greatest financial return for the money invested to achieve it. Each of these methods uses slightly different approaches and answers different economic evaluation research questions. They also have different advantages and disadvantages. In spite of these differences and limitations, they can provide important insight to decision-makers on the economic outputs of health interventions compared to the economic inputs required to produce them.

References

Alemi, F., Haack, M., Holifield, L., Claudio, Y., & Haqqi, K. (2002). Unit cost of counseling and patients length of stay in a residential drug treatment setting. *Journal of Mental Health Policy and Economics, 5,* 103–107.

Andresen, M. A., & Boyd, N. (2010). A cost-benefit and cost-effectiveness analysis of Vancouver's supervised injection facility. *International Journal of Drug Policy, 21*(1), 70–76.

Baltussen, R., & Niessen, L. (2006). Priority setting of health interventions: The need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation, 4,* 14.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2011). *Cost-benefit analysis: Concepts and practice* (4th ed.). Boston, MA: Pearson Eduction Inc.

Brouwer, W. B. F., van Exel, N. J. A., Baltussen, R. M. P. M., & Rutten, F. F. H. (2006). A dollar is a dollar is a dollar—Or is it? *Value in Health, 9,* 341–347.

Byford, S., Torgerson, D. J., & Raftery, J. (2000). Cost of illness studies. *British Medical Journal, 320,* 1335.

Cartwright, W. S. (2008). Economic costs of drug abuse: Financial, cost of illness and services. *Journal of Substance Abuse Treatment, 34,* 224–233.

Clabaugh, G., & Ward, M. M. (2008). Cost-of-illness studies in the United States: A systematic review of methodologies used for direct cost. *Value in Health, 11,* 13–21.

Currie, G., Kerfoot, K. D., Donaldson, C., & Macarthur, C. (2000). Are cost of injury studies useful? *Injury Prevention, 6,* 175–176.

Diamond, G. A., & Kaul, S. (2009). Cost, effectiveness, and cost-effectiveness. *Circulation: Cardiovascular Quality and Outcomes*, *2*, 49–54.

Dowdy, D., Rodriguez, R., Hare, B., & Kaplan, B. (2011). Cost-effectiveness of targeted human immunodeficiency virus screening in an urban emergency room. *Academic Emergency Medicine*, *18*(7).

Drummond, M. (1992). Cost-of-illness studies: A major headache? *Pharmacoeconomics, 2,* 1–4.

Fleming, M. F., Mundt, M. P., French, M. T., Manwell, L. B., Stauffacher, E. A., & Barry, K. L. (2002). Brief physician advice for problem drinkers: Long-term efficacy and benefit-cost analysis. *Alcoholism, Clinical and Experimental Research, 26*(1), 36–43.

French, M. T. (1995). Economic evaluation of drug abuse treatment programs: Methodology and findings. *American Journal of Drug and Alcohol Abuse, 21*(1), 111–135.

Gafni, A. (1994). The standard gamble method: What is being measured and how it is interpreted. *Health Service Research*, *29*(2), 207–224.

Gentilello, L., Ebel, B., Wickizer, T., Salkever, D., & Rivara, F. (2005). Alcohol interventions for trauma patients treated in emergency departments and hospitals: A cost benefit analysis. *Annals of Surgery, 241*(4), 541–550.

Gold, M., Siegel, J., Russell, L., & Weinstein, M. (1996). *Cost-effectiveness in health and medicine*. New York, NY: Oxford University Press.

Hansson, S. O. (2007). Philosophical problems in cost-benefit analysis. *Economics and Philosophy, 23,* 163–183.

Hennessy, C., Moriarty, D., Zack, M., Scherr, P., & Brackbill, R. (1994). Measuring health-related quality of life for public health surveillance. *Public Health Reports, 109*(5), 665–672.

Holder, H., Longabaugh, R., Miller, W. R., & Rubonis, A. V. (1991). The cost effectiveness of treatment for alcoholism: A first approximation. *Journal of Studies on Alcohol, 52*(6), 517–540.

Kleiman, M. A. (1999). "Economic cost" measurements, damage minimization and drug abuse control policy. *Addiction, 94,* 638–641.

Mankiw, N. G. (2012). *Principles of economics* (6th ed.). Boston, MA: Cengage Learning.

McCabe, C. (2009). *What is cost utility analysis?* Hayward Medical Communications. Available at: http://www.whatisseries.co.uk/christopher-mccabe/

Murphy, S., Sales, P., & Averill, S. (2015). Ethnographic studies of drugs in community. In H. H. Brownstein (Ed.), *The handbook of drugs in society*. West Sussex, UK: Wiley.

Neumann, P., Cohen, J., & Weinstein, M. (2014). Updating cost-effectiveness—The curious resilience of the $50,000-per-QALY threshold. *New England Journal of Medicine, 371,* 796–797.

Plotnick, R. D. (1994). Applying benefit-cost analysis to substance use prevention programs. *International Journal of the Additctions, 29*(3), 339–359.

Quanbeck, A., Lang, K., Enami, K., & Brown, R. L. (2010). A cost-benefit analysis of Wisconsin's screening, brief intervention, and referral to treatment program: Adding the employer's perspective. *Wisconsin Medical Journal, 109*(1), 9–14.

Robinson, R. (1993a). Economic evaluation and health care: What does it mean? *British Medical Journal, 307,* 670–673.

Robinson, R. (1993b). Cost-benefit analysis. *British Medical Journal, 307,* 924–926.

Rosen, H. S., & Gayer, T. (2008). *Public finance* (8th ed.). Boston, MA: McGraw-Hill Irwin.

Segel, J. (2001). *Cost-of-illness studies: A primer.* Research Triangle Park, NC: RTI International, RTI-UNC Center of Excellence in Health Promotion Economics.

Shiell, A., Gerard, K., & Donaldson, C. (1987). Cost of illness studies: An aid to decision-making? *Health Policy, 8,* 317–323.

Trumbull, W. N. (1990). Who has standing in cost-benefit analysis? *Journal of Policy Analysis and Management, 9,* 201–218.

Vanhook, P. (2007) Cost-utility analysis: A method of quantifying the value of registered nurses. *Online Journal of Issues in Nursing*, *12*(3).

Whittington, D., & MacRae, D. (1986). The issue of standing in cost-benefit analysis. *Journal of Policy Analysis and Management, 5,* 665–682.

Wiseman, V., & Mooney, G. (1998). Burden of illness estimates for priority setting: A debate revisited. *Health Policy, 43,* 243–251.

Evaluation of Substance Abuse Prevention and Treatment Programs

20

Peggy Stephens, Zili Sloboda, and Deric Kenne

20.1 Introduction

This chapter will focus on evaluation approaches to adolescent substance use prevention in schools and communities. For clarification purposes, the term "substance" refers specifically to any chemical substance, either in its natural or man-made form, which alters biological structures or functioning when administered and absorbed. Such substances that affect feelings, perceptions, thought processes, and/or behavior fall under the rubric of psychoactive substances, which exert their effects by altering the function of the nervous system. Psychoactive substances include legal substances such as alcohol, tobacco, and certain prescription drugs, while examples of illegal substances include hashish/cannabis, heroin, and methamphetamines. Furthermore, the term *use* is employed, as opposed to abuse, because any use of these substances by adolescents is problematic from a public health (PH) perspective; as the goal of most PH interventions is to prevent behaviors that lead to problematic health outcomes. Before delving into how to approach the evaluation of programs designed to prevent the use of psychoactive substances (herein referred to as substance or substances), current approaches to substance use prevention are summarized briefly.

In the United States, prevention interventions have been implemented within schools, the workplace, and in communities through families and prevention policies, with the goal of preventing (or reducing) substance use by children and adolescents. Prior to the late 1980s, prevention intervention programs had little empirical evidence of achieving these goals, but a growing body of evidence over the past three decades has provided public health interventionists with information to guide them in the selection and implementation of evidence-based or promising programs that can achieve substance use prevention goals in specific populations and communities. This accumulation of evidence comes by and large from the efforts of researchers and evaluators who have systematically examined the extent to which substance use prevention programs (1) achieve reductions in substance use, (2) are effective under diverse levels of implementation and in different

P. Stephens (✉)
Department of Social and Behavioral Sciences, College of Public Health, Kent State University, Kent, OH, USA
e-mail: msteph16@kent.edu

Z. Sloboda
Applied Prevention Science International, Inc., Ontario, OH, USA
e-mail: Zili.sloboda@apsintl.org

D. Kenne
Department of Health Policy and Management, College of Public Health, Kent State University, Kent, OH, USA
e-mail: dkenne@kent.edu

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3_20

contexts, and (3) are comparatively efficient for use by schools and communities (Botvin and Griffin 2003; Schinke et al. 1991).

This work has contributed to the formation of the emerging field of prevention science and the establishment of the Society for Prevention Research, both in the United States and the European Union. Prevention science can be defined by its contributing disciplines, which include epidemiology, psychology, sociology, economics, neurobiology, and genetics, each with their associated theories and research methodologies. It is by way of prevention science that a better understanding of both etiology and the natural history of substance use has been realized, as well as proper design and evaluation of prevention interventions.

Another major contribution to the field was the work of Hawkins and his colleagues, published in 1992, which summarized the findings from epidemiological and etiological studies highlighting and organizing factors that were found to be associated with the initiation of substance use (Hawkins et al. 1992). These factors fell into two broad categories, comprised of those related to context or the environment, and those related to intrapersonal or individual deficiencies. This work had a major impact on the field of substance use prevention, particularly focusing on the issue of susceptibility and risk. However, the work did not address the mechanisms involved with the onset of substance use, nor with the progression from use to abuse and dependence. Findings from recent etiological studies focusing on problem behaviors are beginning to provide more information regarding the mechanisms underlying the initiation of substance using behaviors, suggesting that the interaction or interface between the individual and his/her micro-level (family, school, peers, workplace) and macro-level environments (neighborhood, community, nations) shapes the embrace of prosocial or anti-social attitudes and behaviors (O'Connor and Rutter 1996; Fishbein and Ridenour 2013; Sloboda 2015a, b). Vulnerability and increased susceptibility may arise due to failure to meet developmental benchmarks, challenges related to adolescent developmental issues, or to negative life events. However, vulnerability alone is not sufficient to put individuals on a negative life trajectory. The environment, particularly the micro-level environment, which includes family, school, peers, the workplace, and faith-based organizations, is a key ingredient in this process. For example, vulnerable children with parents who evidence good parenting skills are more likely to avoid engagement in problem behaviors than those children whose parents are neglectful and/or non-supportive (Kumpfer and Alvarado 2003).

In addition, related behavioral theories, such as the Theory of Planned Behavior and its variations (Ajzen 2002), Self-Efficacy Theory (Bandura 1997), the Theory of Triadic Influence (Flay et al. 2009), and theories of learning (Piaget 1973; Bloom 1956) have contributed to the design and development of prevention interventions. The design of prevention interventions and their theories or conceptual foundations are important for framing the evaluation approach, not only including what measurements are most relevant but also the sampling plan, research design, and analytic methods to be used.

20.2 Evidence-Based Substance Use Prevention Interventions and Policies

Currently, a number of international and national organizations, including the United Nations, the European Monitoring Centre for Drug Addiction and Drugs, the U.S. National Institute on Drug Abuse and its sister organizations, the Substance Abuse and Mental Health Services Administration and the Office of Juvenile Justice and Delinquency Prevention, have published guides on both evidence-based substance use prevention interventions and policies (UNODC 2013). These guides indicate a range of evidence supporting substance use prevention approaches for infants and children at risk for negative social and health outcomes. Prevention interventions include both behavioral and policy approaches that target families, schools, the workplace, and the community. There is an accumulative

evidence base supporting the implementation of comprehensive, multilevel prevention approaches that provide programming at least during middle and high school (and preferably early on in elementary school), including, but not limited to, media and/or policy components that target parents and the community at large (Carson et al. 2011; Foxcroft and Tsertsvadze 2011; Jackson et al. 2012).

Three programming approaches are generally recommended: (1) universal programs, which are designed to prevent the onset of use for the broader population of adolescents, (2) selective programs, which target those at higher risk for use, and (3) indicated programs for adolescents who show the greatest risk of use (Bukoski 2003). Due to space limitations, this chapter will focus on universal programs; however, the evaluation principles and procedures outlined apply to any program, at any level, and for any population. The nature of substance use epidemiology in any community where risk status varies across the population warrants not only all levels of prevention programming, but also the integration of prevention and treatment services (Sloboda 2015a, b).

20.3 The School as a Prevention Setting

The majority of research conducted on substance use prevention interventions has focused on the school. Schools are ideal settings for prevention programming as they reach the largest number of children and adolescents. Additionally, schools are considered key socialization agents in any community and society in preparing children to assume their roles as productive adults (Garbarino 1978; Petras and Sloboda 2014); teaching the cognitive, social, and life skills necessary to survive. Schools are natural sites for prevention interventions, providing at least three opportunities for the introduction of prevention interventions, which include: (1) school culture; that is, the norms, beliefs, expectancies, and school bonding, which connects the individual to the school experience and community, (2) school policy or social control, the most common approach to disciplinary policies and procedures, and (3) classroom curriculum, or packaged programs (Sloboda 2009).

The most prevalent type of school-based substance use prevention intervention is the classroom curriculum. Evidence-based curricula are generally tailored to specific problematic substances in the target population. Ideally, they also include content acknowledging social influences on substance use, normative beliefs regarding peer substance use, resistance skills that enhance students' ability to both recognize and avoid situations where drugs may be available, and to act assertively with a commitment not to use drugs. Moreover, the format for programming should be interactive, with students actively engaging in discussion, role play, and other activities. Finally, programming should be developmentally and culturally appropriate to the target population and administered on an ongoing basis with booster sessions at regular intervals (Bukoski 2003; National Institute on Drug Abuse 2003; UNODC 2013).

20.4 Program Evaluation

Program evaluation, in the broadest sense, is the systematic assessment of the extent to which programs achieve stated goals, under what circumstances those goals are achieved, and to what extent interventions are cost efficient in achieving those goals. Evaluation utilizes qualitative and quantitative research methods, but is different from basic research in that the goal is not only knowledge acquisition, but knowledge acquisition with the purpose of direct application to improving public health interventions.

There are two broad types of evaluation: formative and summative (Weiss 1998). Formative evaluation, or process evaluation, is conducted for the purpose of providing immediate feedback and opportunity for adjustments in either the program itself or the way it is delivered. Of particular interest is the extent to which programs are able to be implemented in the real world (feasibility) and/or put into practice in the

manner in which they were designed (implementation fidelity). While it is usually utilized in the early stages of program development, it may also include an examination of immediate outcomes or mediators of the intervention (knowledge, skills, perceptions/attitudes) with the intent of making programmatic adjustments (content implementation or emphasis) to improve those, and ultimately more distal outcomes. For example, in an adolescent substance use prevention program designed to target social skills, with the ultimate goal of increasing students' ability to stand up to peer pressure to initiate alcohol use, evaluators may want to examine whether or not the curriculum has changed those skills early on, immediately after the students are exposed to the curriculum. If the social skills show no change immediately after the program, it would then make sense to make some adjustments to the program because without an increase in social skills, there would be no expected change in the ability of students to stand up to peer pressure to use alcohol.

Formative evaluation employs, qualitative, quantitative, and mixed methods designs (Castro et al. 2014) to address questions such as: (1) Do we have the resources to deliver the program to the target population? (2) Did the training increase instructors' confidence in their ability to deliver the content of the program? (3) How many program sessions were attended by the target population? (4) Did the program recipients engage in the activities offered by the program? and (5) Was the program content delivered uniformly by all the instructors?

Summative evaluation, also known as outcome evaluation, focuses on assessing the changes produced by the program in terms of immediate, intermediate, and long-term goals. In substance use prevention, effective programs target or try to change decision-making processes and intentions to use substances. Hypothetically, by changing these immediate outcomes, or by improving decision-making skills within a prosocial context and the intermediate outcomes (e.g., adoption of prosocial attitudes and intending to abstain from substance use), there will be a consequent change in the outcome of interest: substance use. Such an evaluation approach examines whether or not the program has changed immediate targets, followed by intermediate and ultimate program targets. Impact evaluation goes further and examines the effect of a set of program activities on the outcome of interest (Mohr 1995). Summative evaluation relies heavily on quantitative research methods, including experimental and quasi-experimental designs, and statistical techniques such as mediational analysis to test causal relationships and calculate program effect sizes for specific constructs or behaviors the program is expected to change (McKinnon et al. 2007). Summative evaluation also includes analyses conducted to determine whether or not the intervention does its job efficiently. Efficiency refers to the balance between the resources required to implement the program and the value associated with the effects of the program; efficiency is judged by way of economic evaluation (see Oglesby and Birmingham, this volume). As economic evaluation is addressed elsewhere in this volume, it will not be addressed here. It is worth noting, however, that if a program does not show an effect on the targeted outcomes (it is not effective), there would be no reason to move forward in conducting an economic evaluation. Moreover, it is important to consider cost, even with effective programs.

20.4.1 Why Evaluate Prevention Programs?

Program evaluation provides the information needed to make decisions regarding which programs to select for implementation, whether a program should be continued or discontinued, and if continued, with or without changes and improvements. Evaluation should also be viewed as an ongoing process, and utilized to inform decision-makers, participants, and other stakeholders on program strengths, weaknesses, and future directions (CDC 2011).

The CDC identifies four sets of standards for evaluation activities: (1) utility, (2) feasibility, (3) propriety, and (4) accuracy. These standards may be applied to the evaluation of any public

health intervention, including substance use prevention programs. The utility standard calls for evaluation activities that meet the needs of participating stakeholders. Ideally, those needs are determined by an assessment of the prevalence of substance use in the target population, the social and environmental risks, the protective factors associated with substance use, and the social and economic resources available to address those factors (Arthur and Blitz 2000). Any evaluation approach should focus on providing information to assist the program developers, participants, and other interested parties in making decisions about program improvements, expansion to larger populations or different contexts, or to make comparisons of the target program to other potential interventions. Feasibility standards include whether or not the evaluation procedures can actually be carried out, are justified in terms of the resources expended, and are politically realistic regarding stakeholder support and commitment to the program. The propriety standard demands that evaluation activities be undertaken in ways that are not only legal, but ethical, with a service orientation and transparency of procedures and findings. Finally, accuracy standards require that the evaluator utilize appropriate research methodology, including data collection and interpretation, to draw valid conclusions and unbiased reporting of the findings of the study (CDC 2011).

20.4.2 Who Does Evaluation?

Program evaluation may be conducted internally by those who design or implement the program, or externally by someone who has no direct stake in the success or failure of the program being assessed. External evaluators may be academics, experienced researchers, or evaluators with knowledge and experience in conducting evaluations. They can provide expertise in focusing the evaluation, matching the data collection and analyses to anticipated outcomes, constructing measures for inputs, activities, outputs, and outcomes (see next section), analyzing the data, and summarizing the results. The former is desirable in that those involved in the program selection, development, and delivery should, as a form of organizational accountability, have some plan to continuously monitor and assess each phase of this process.

Those involved in administering a program will understand the problem the program is designed to address, the target population, and its activities and goals better than someone outside the organization. Consequently, internal evaluators will have insight into the nuances of program administration, as well as receptivity, which may not be evident to an outsider. Internal evaluations are also generally less costly than external evaluations because of the lower costs for staff or personnel to plan the evaluation, collect and analyze data, and write and disseminate reports.

In practice, however, many organizations do not conduct any type of program evaluation. In fact, most school-based substance use prevention programs are developed locally, and little is known about the extent to which these programs have been tested for effectiveness (Kumar et al. 2013). One reason for this lack of internal assessment may be that those involved in the program do not always have the skills necessary to conduct a rigorous evaluation. Another major weakness of relying on internal evaluation is the potential bias posed by conducting an evaluation on a program in which the organization, or individuals within the organization, are vested. By contracting with external evaluators, this bias is reduced. Furthermore, professional evaluators can assist the organization in developing data collection and monitoring infrastructure for continuing the evaluation process after the main study is completed. The choice of internal versus external evaluation is not necessarily an either/or choice. There is also the option of a hybrid evaluation, in which program staff conducts the evaluation with expert consultation.

20.5 Planning the Evaluation

Program evaluation may be initiated by individuals who have a role in developing, administering, funding, or participating in the program, or

anyone who has some interest in determining the value of the program. Once the decision to evaluate has been made, the first order of business is to identify relevant stakeholders. Stakeholders are those individuals, groups, or organizations who have some involvement or interest in the program. Stakeholders include those targeted by the program (participants), those who are involved in administering the program, and those who have an interest in the expected outcomes. For example, stakeholders in a school-based substance use prevention program might include school administrators, teachers, prevention specialists, parents, students, local public health officials, and law enforcement and substance abuse treatment agencies. Gatekeepers are stakeholders who can deny access to those who might be conducting an evaluation.

20.5.1 Defining Program Goals and Processes

While it may seem obvious, probably the most important process in planning an evaluation is to come to an agreement upon a description of the program or intervention. The program may be newly developed (and not yet incorporated into practice) or have been in place for some time, but neither scenario guarantees that the program actually has clear and/or agreed upon definitions of the target population, content and activities, or the individuals or organizations responsible for delivering the program. To undertake this activity, a key group of stakeholders who are directly involved in the program should be assembled with the purpose of producing a detailed program description, beginning with the problem the program is expected to address (e.g., drug abuse prevention, alcohol abuse treatment), and followed by explicit program goals. After these critical components have been agreed upon, the group can move on to describe the activities undertaken as well as the resources and personnel responsible for accomplishing each activity. The next section describes two tools that facilitate this process.

Program theory is a conceptual model of how the content and activities of the intervention will change the program targets. Even if there is no explicit behavioral theory underlying the program, all programs have a set of implicit assumptions that if we do these things, then we expect something to happen (McLaughlin and Jordan 2004). For example, if we change students' beliefs about the harmfulness of alcohol use, then they will make a commitment to avoid drinking alcohol. A visual representation of a hypothetical school-based alcohol use prevention program is presented in Fig. 20.1.

In the hypothetical school-based alcohol prevention example presented above, the program theory is explicated and proposes the following: If students are exposed to ten, one-hour sessions

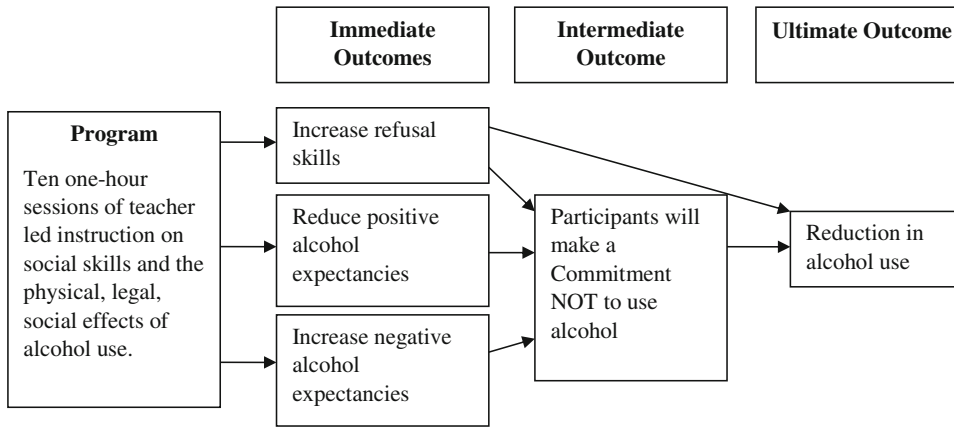

Fig. 20.1 Program theory for alcohol prevention program

comprising the alcohol intervention program, then they will increase their ability to refuse an offer of alcohol, reduce their expectations that alcohol will have a positive effect on them, and increase their perceptions that alcohol will have a negative effect on them, and as such, the immediate outcomes will be achieved. If these three attributes are changed, then students will make a commitment not to use alcohol (intermediate outcome), and if they make that commitment, then they will not use alcohol (ultimate outcome). One important function of the program theory is that it identifies the constructs that the program is hypothesized to change directly (refusal skills and alcohol expectancies) and also those characteristics that are expected to be affected only indirectly, as a result of changes in the direct program targets (commitment not to use and actual behavioral use of alcohol). This identification of the immediate, intermediate, and ultimate outcomes (or indirect and direct effects on the ultimate goal of alcohol use) is useful for planning the evaluation because it directs the evaluators to what constructs should be measured and enables them to plan and conduct analyses to assess where the program worked and where it may have failed (Bickman 1987). For example, if the evaluation shows the program did not change alcohol use, evaluators would work backwards in the theory to examine whether or not the program's constructs need to be changed in order for alcohol use to change.

A logic model is a useful tool in explicitly stating these cause and effect relationships and describing, in concrete terms, the processes that must be put in place for those changes to take place. It breaks down the program into its component parts and facilitates evaluation planning for the process of implementing the program, as well as for determining the program's effect on target outcomes. Any program can be defined by its inputs, activities, outputs, and outcomes. The logic model organizes these components in one of several forms, such as a flow chart or organizational grid, with the purpose of displaying how the goals of the program are linked back to the activities and resources used to achieve those goals. Inputs may be thought of as the resources put into a program to perform activities that are necessary to achieve the program goals, while outputs are the immediate products of the inputs and activities. The outputs from might be to expose 100 students to the alcohol curriculum. The results of this set of inputs, activities, and outputs would be to change the immediate, intermediate, and ultimate outcomes targeted by the program. In an evaluation context, inputs, activities, and outputs are examined through process analysis while outcomes are appraised by way of outcome analysis.

A logic model organizes these components for each specific program goal. A complete logic model is too large to include in this chapter, but included is a single segment of a logic model for the program theory shown in Table 20.1. A complete logic model would include one segment for each of the goals and objectives in the program. The W. K. Kellogg Foundation (2004a, b) is an excellent resource for developing and utilizing logic models in conducting program evaluation; these resources may be found at www.wkkf.org.

20.5.2 Focusing the Evaluation

The resources available to evaluate program components are usually a limitation on the extent to which those components can be examined. For this reason, program implementers and evaluators need to come to an agreement on which components of the program will be examined. For example, will the resources only support an outcome analysis or can we examine both the process of implementing the program along with the outcomes (goals) of the program? Both program theory and logic models can assist in this task. When administrators and evaluators understand what the specific goals and procedures for achieving those goals are, they can prioritize which goals and processes can, and should, be examined given the resources available for the evaluation study. Every goal and process that is targeted for assessment must be measured, analyzed, and reported. Developing, collecting, entering, and analyzing data are

Table 20.1 Example segment of logic model for alcohol program

Goal	Inputs	Activities	Outputs	Immediate outcomes	Intermediate outcome	Ultimate outcome
Deliver ten, one-hour sessions of prevention curriculum to 50 6th graders and assignment of 50 students to the control condition (no program)	Administrative assistant time for recruiting students and assigning to intervention or control group. Space for class sessions Health teacher instructor time for ten groups for ten sessions	Administrative Assistant identifies students, contacts parents for consent, orders instructor and student materials, identifies and contracts for space, schedules class administration sessions Instructor administers ten one-hour sessions of curriculum	Space for ten, one-hour class sessions Completion of the ten sessions by 50 students and 50 students identified as control group. One set of instructor materials and 50 sets of student materials	100% of the intervention students will attend 10 sessions of the classes	Intervention students will demonstrate knowledge of social skills and the physical, social, legal effects of alcohol	Ninety-five percent of intervention students will make a public commitment not to use alcohol until legal age

time-intensive and require resources to be completed. If the program administrators wish to examine the implementation fidelity of the program in addition to the outcomes, more data collection is required, in addition to the handling and analyses of those data.

Program theory and logic models assist those who are conducting the evaluation in determining the critical components of the program in the causal chain leading to the program goals, as well as the budgetary requirements for examining each process. The overarching questions for focusing the evaluation are, "What are the most important goals the program is proposed to achieve?" and "Which inputs, activities, outputs, and outcomes are critical for realizing those high priority goals?"

20.5.3 Evaluation Design

Selection of the evaluation design begins with the questions the evaluator is charged with answering. The question of whether or not the program produced a change in the outcomes of interest (outcome evaluation) leads the evaluator to choose a causal design (experimental or quasi-experimental design). The question of whether or not the program was implemented as intended (process evaluation) leads the evaluator to an observational or descriptive design. It is beyond the scope of this chapter to cover all the types of research designs available within these two broad categories, and other chapters in this volume address causal and descriptive research designs.

Causal research designs are utilized to test hypotheses about cause and effect relationships. In an evaluation context, the cause is the program and the effect is the change of the behavior, or health outcome, that the program is designed to change. For example, the alcohol program is proposed to cause change in refusal skills and expectancies related to alcohol use. This can also be thought of as the cause being an independent variable (X) and each outcome a dependent variable (Y). The most valid research design to test cause and effect relationships are experiments where units of interest (those being exposed to the program) are randomly assigned

to either an intervention group, which receives the program, or a control group that does not receive the program (see Swartz, this volume). In causal research designs, internal validity is the confidence the evaluator/researcher has that whatever change is observed in the outcome (*Y*) is due to the program (*X*) and not some other spurious or confounding influence on the outcome (see Peron et al., this volume).

For example, how can one be confident that exposure to the ten alcohol prevention sessions increased refusal skills, as opposed to some other event the students had experienced between the pre-test and post-test? If we evaluated a single group of students who were exposed to the program, there would be several reasons why we might doubt that the program was the only influence on refusal skills. It is possible, for example, that students may have developed better refusal skills simply because they had matured during the subsequent time period as would be expected in their normal course of development. This is known as maturation, and is only one of several possible alternative explanations as to why it appears that a single group of students, exposed to an intervention, showed change in a targeted outcome. Possible alternative explanations for a change in *Y* when a program is implemented are referred to as threats to internal validity. In addition to maturation, other common threats to internal validity include selection bias (the possibility that the characteristics of participants contributed to the observed change as opposed to the program itself), history (unanticipated events or experiences that may occur while the intervention is in progress that change participant knowledge and/or attitudes), testing (subjects "learning" from prior assessments in the evaluation), instrumentation (changes in the characteristics of the measurement instruments utilized), statistical regression (a tendency for subjects selected on the bases of extreme scores to regress towards the mean on subsequent tests), and mortality (Cook and Campbell 1979).

Experimental designs include critical components that minimize or reduce these threats. The randomization of units, such as students or schools, to either a control or the treatment/intervention group offers protection against invalid conclusions regarding the effect of the program. The control group that is not exposed to the program provides the evaluator with an estimate referred to as the counterfactual, or a description of the group with respect to the outcome variables if there was no intervention; representing the "null" condition (what could be expected if it was decided, for example, to do nothing about adolescent alcohol use during period of time of the intervention). Randomizing units to either control or intervention groups insures that at the starting point (pre-test or pre-intervention) of the evaluation, students in the two groups are, on average, equivalent on their levels of not only the outcomes of interest, but also any other characteristics (sex, race, age, etc.) that may influence the effects of the program on those outcomes. The equivalence produced by randomization reduces the possibility that any difference (effect) seen between the control and intervention group on the outcome(s) is due to one of the threats to internal validity rather than to the intervention program itself.

Another necessary component of an experiment is at least one post-test measure of the outcome variable(s) to measure this effect (a pre-test and multiple post-tests may also be added to the basic single post-test design). The addition of a pre-test measure of the outcome variable(s) provides the evaluator with an estimate of where the subjects are on the level of the dependent variable before the program, and when compared to the post-test scores, how much change occurred in each group between the pre-test and the post-test. A pre-test also provides measures with which to check the randomization process; by including demographic variables along with measures of the outcomes at the pre-test, the evaluator can compare the control and treatment/intervention group on these characteristics to insure that the randomization process produces equivalence between the groups.

Oftentimes, it is not feasible to have a control group, or otherwise randomized groups. Under these circumstances, the evaluator will need to compromise on the experimental design and

move to a quasi-experimental design, perhaps using a nonequivalent comparison group or no comparison group at all. All quasi-experiments expose the evaluation findings to threats of internal validity, the seriousness of which depends upon which alternative design is selected (Shadish et al. 2002).[1]

While experimental designs, and to some extent quasi-experimental designs, may have good internal reliability, they do not necessarily have good external validity, or the ability to generalize findings to the target population. External validity is achieved through the use of random probability sampling from the target population. A poorly designed or executed sampling strategy will most likely result in a selection bias, resulting from poor coverage of the population or biased participant response rates during recruitment, consent, and/or data collection (Brown 2006). However, experiments can have good external validity (in addition to internal validity) if participants are randomly sampled from a well-defined population before being randomized to the intervention or control groups.

In contrast to experimental studies, descriptive or observational research designs are utilized to answer questions regarding the extent to which the program is implemented as planned. Descriptive designs involve the collection of observations and data with the intent of describing the inputs, activities, and outputs designated as critical for program success. Descriptive studies do not propose causation; rather, they provide evidence of the link between what is proposed and what actually happens during program administration. Both quantitative and qualitative approaches to data collection and analysis may be useful for studying program processes, for which approach selection depends upon largely on the processes being examined. Baranowski and Stables (2000) identify critical components of process evaluation to be examined

> Thus, a minimum useful number of components of process evaluation appear to be recruitment and maintenance of participants, context within which the program functions, resources available to the program and the participants, implementation of the program, reach of materials into (or receipt by) the target group, barriers to implementing the program, initial use of program activities, continued use of program specified activities, and contamination of treatment and control groups.

Sources of data reflecting these components include administrative files, secondary data on neighborhood or community characteristics, budget and personnel audits, observations of program delivery, records of attendance for program sessions, and personal interviews or focus groups with program staff, participants, and administrators. The question of "what data should be collected?" is answered by identifying the priorities the stakeholder workgroup has set prior to the evaluation, with respect to program aspects that should and can be evaluated given the program goals and resources available. For example, stakeholders in an evaluation of a community-based prevention program may be primarily interested in outcomes targeted by the program, but a second priority may be to describe the extent to which the resources devoted to the program are actually directed at the program activities and not other programs or projects within the organization. Interviews with program administrators and instructors, as well the collection of administrative data from workload reports and budget expenditures, may be utilized for this purpose. The resources to collect these data may preclude data collection to measure other program processes or influences, but to the program administrators and stakeholders, the tradeoff is necessary to understand if the program is being implemented efficiently.

20.5.4 Ethical Considerations

Evaluation work, which involves working with and collecting data from humans, inherently brings ethical considerations in both how subjects are treated and in regard to the nature of the data collected. A number of ethical challenges are

[1]For a thorough explication of the utilization of experimental and quasi-experimental designs for outcome analyses please refer to Mohr (1995).

evident, including tensions between personal privacy and public accountability, informed consent, cultural competency, and potential conflicts of interest that may arise within relationships between evaluators and program managers/distributors (Gorman and Conde 2007; Pandiani et al. 1998; Rodi and Paget 2007; Schwandt 2007; Hatry et al. 2010). Additionally, there may be unintended ethical threats related to evaluation practice—particularly linked to data collection and evaluation of results—associated with multitiered and international designs (Rodi and Paget 2007; Schwandt 2007).[2] The very nature of evaluation science itself contributes to these ethical challenges. Simons (2006) notes that the public role of the evaluator is to judge the essential values and merits of programs and to disseminate results. In some cases world views can clash in this judgement. Additionally, evaluation is an inherently political endeavor; fraught with potential conflicts between stakeholders, some with vested financial interests in the programs under assessment. While these issues are true of evaluation generally, they are of particular concern in evaluations of substance abuse prevention and treatment programs, as they often involve at-risk populations and highly sensitive data.

As a result, evaluators of substance abuse prevention programs must behave ethically in general, but also go beyond general expectations to address special concerns, such as the collection of data that presents risk for participant embarrassment or legal liability. Weiss (1998) points out that evaluators must be diligent in identifying and addressing ethical concerns in all areas of the evaluation, beginning with the planning phases of the study and continuing through the reporting of results. She condenses these responsibilities into two rules: "Do not harm the people studied, and do not distort the data" (1998). General rules for conducting ethical evaluations include honesty, transparency with all involved in the study, insuring confidentiality and informed consent, and identifying any biases or conflict of interests that may influence the activities or conclusions of the evaluation.

There are also legal considerations for evaluators depending upon the goals and funding for the project being examined. In general, all evaluators and researchers should consider the three principles that guide the conduct of research with human subject included in the Belmont Report (1979): (1) respect for person, (2) beneficence, and (3) justice.[3] However, research has more stringent specific requirements than "practice," including the approval of the study by an internal review board. Historically, the distinguishing of research from practice lies in the goal of the study, as research is conducted to contribute to knowledge that may not directly benefit those participating in the study. Evaluation, which generally falls under the auspices of practice, is meant to produce information or knowledge to improve a specific program (Belmont Report 1979).[4]

The first principle cited in the Belmont Report is that of respect for persons (1979). This principle demands that subject participation in any part of the evaluation should be voluntary and

[2]These threats are often associated with conflicting evaluation structures and contracting arrangements that exist across subdivisions with separate administrative and organizational structures (Rodi and Paget 2007). Different values and norms may also need to be negotiated by the evaluator, especially in international designs.

[3]On July 12, 1974, the National Research Act (Pub. L. 93-348) was signed into law, creating the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. One of the charges to the Commission was to identify the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects and to develop guidelines which should be followed to assure that such research is conducted in accordance with those principles. These principles are summarized in the Belmont Report, available at http://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/.

[4]Evaluation often falls under the classification of "practice," However, if the evaluator is planning to utilize the data collected to contribute to knowledge beyond only the specific program under study, he/she should review the Code of Federal Regulations: 45 CFR 46 (Dept. of Health and Human Services 2009), particularly if the source of funding for the evaluation is federally-based. Furthermore, state and local regulations may vary regarding evaluations conducted in schools or communities; it is the responsibility of the evaluator to be informed and in compliance with these regulations.

obtained without coercion, such as the use of incentives or rewards that would be difficult to refuse. Subjects must be able to give informed consent for their participation, which means that the evaluator must disclose the purpose of their involvement in the evaluation, any potential risks and benefits, the length of project involvement, and all that involvement entails (completing surveys, interviews, access to school or legal records, etc.).

A second ethical principle is that of beneficence. Beneficence requires that the evaluator balance any potential risks to the participants with potential benefits that they may receive from participation. As such, the risks should be minimal and the benefits should justify any potential risks. The balance implies that the evaluator "does no harm" to participants—even if the rewards or benefits from participation are large. Typically, participants in substance use prevention evaluations are asked to disclose the use of illegal substances and other risk factors, including any illegal behaviors of both the participant and his/her family or friends. Another potential problem is that legal authorities may learn of these disclosures, with participants potentially suffering social or legal sanctions as a result. It is the responsibility of the evaluator to assess the risk of unintended disclosure and to plan for procedures that minimize those risks. For example, the evaluation plan should include documented procedures for the collection and storage of data to maintain anonymity or confidentiality, along with training for anyone who will be involved in the data collection.

Finally, the justice principle requires that no individual or group of participants incurs greater risk, or reaps more benefit of an evaluation, than other participants. For example, it would be unethical under the justice principle to provide a program to only those schools or communities that can afford to implement the program, while relegating the poorer schools or communities to comparison conditions because they lack the resources to implement programming.[5]

[5]The same three principles apply to the handling and reporting of data. Evaluators should insure that data are collected in a manner that offers privacy to the individuals from whom the data are collected, that data collection procedures from secondary sources or administrative files are explicitly designed and conducted so that files are not exposed to unauthorized access, and data are transported and stored in a secure manner, including the removal of any identifying information (e.g., names, birthdates) from individual level data, in addition to other characteristics that may identify groups or communities in community studies. If the evaluation is longitudinal, a "key" file should be created to link data collected at different points in time, and should be stored separately from the evaluation data. An extra layer of protection for confidential data can be secured with a certificate of confidentiality, which protects the researcher from having confidential data subpoenaed in court. In the United States, these are typically obtained from the funding office, or through research offices located at most universities.

20.5.5 Measurement Issues: Special Considerations for Evaluation

The identification of constructs to measure and report can only be finalized after the evaluator and stakeholders have decided on the program goals, inputs, activities, and outputs, and on which of these components the evaluation will focus. Concepts are the broader ideas the evaluator wishes to examine (e.g., alcohol use, commitment not to use alcohol, instructor engagement in program activities), and operationalization is how the construct is observed in specific, recordable terms. For example, alcohol use may be operationalized (measured) with a question asking students, "How many days (if any) have you had alcoholic beverages to drink in the past 30 days (more than just a few sips)?" The student would be asked to select the appropriate number of days from a list of numbers ranging from 1 to 30. Three requirements of quality evaluation measures are that they are valid, reliable, and sensitive to change. That is, each measure should be an accurate indicator of the construct it is supposed to be measuring, it should produce comparable results each time it is used in similar situations, and it should show change from the pre-test to post-test(s) if the program has an effect on that construct. Generally, evaluators are well served by using existing measures that have been previously assessed for

validity and reliability rather than creating new measures.[6] The purpose of this section is to highlight the importance of careful consideration of issues involved in measuring constructs in prevention interventions.[7]

While validity is important for measures of all constructs (inputs, activities, outputs, and outcomes), reliability, and sensitivity are critical for outcome measures as the evaluator is looking for change in the immediate, intermediate, and ultimate targeted outcomes. In substance use evaluation, the immediate and intermediate outcomes are referred to as mediators. Mediators are the theoretical constructs which the program proposes to change with the assumption that if these are changed, then substance use will also change. Mediators are commonly measures of attitudes, beliefs, and skill sets, and are often operationalized with responses using a Likert scale. There are unique challenges in measuring mediators and substance use outcomes regarding validity, reliability, and sensitivity. For example, a measure that lacks validity will not provide a reflection of the theoretical model of the program and, consequently, any conclusion as to the accuracy of the model will be questionable. An unreliable measure (one that has an unacceptable level of error) may not show significant change even when the program has changed the construct (Type II error). Similarly, a measure that lacks sensitivity may also lead to a Type II error of failing to reject the null hypothesis of no change/difference when the construct has in fact changed.

Floor or ceiling effects in mediator measures and low prevalence rates for drug use in younger adolescents also present challenges for measuring constructs in prevention interventions. For example, in elementary and middle school most students have negative attitudes toward drugs and, consequently, low rates of use. A question with responses scored by a Likert scale may not be sensitive enough to detect very small changes in attitudes or use, which are the only changes we can expect when most students already exhibit the desired characteristic (e.g., no drug use, negative attitudes towards use). Furthermore, Likert scales, which are ordinal scales, are prone to error. Measures of drug use utilized in epidemiological studies of adolescent risk behaviors typically have response options such as: (0) no use, (1) used on 1 or 2 occasions, (2) used on 3–5 occasions, (3) used on 6–9 occasions, (4) used on 10–19 occasions, (5) used on 20 to 39 occasions, (6) used on 40 or more occasions in the past 30 days, past year, or lifetime (Johnston et al. 2013; CDC 2004).

While these measures are commonly used to evaluate substance use interventions, they pose problems when applied to evaluation studies for the purpose of measuring change in a sample of adolescents. Take, for example, this simple illustration of the potential pitfalls: student A has used marijuana 39 times in the past year when asked to complete the pre-test survey and correctly selects the response labeled "20–39 occasions." The same student has used marijuana 21 times in the year preceding the post-test and again, correctly selects the response labeled "20–39 occasions." Will the evaluator detect any change in student A's marijuana use? The answer is "no," even though this student has obviously reduced his/her drug use during that time period. On the other hand, student B, who indicted "no use" at pre-test, selects the response labeled "used on 1 or 2 occasions" at post-test, indicating a change in the direction of increased use. In this example, the evaluator has not detected change in the expected direction (decrease in use) while detecting change in the opposite direction (increase in use), concluding that the program had the opposite effect as expected. In fact, the program did have an effect in the expected direction, and the effect was quite large. If this pattern held for the entire sample of adolescents, the evaluator would come to the wrong conclusion; the correct

[6]A review of existing literature should produce a variety of measures of substance use risk and protective factors and behavioral use measures to test program theory (Arthur et al. 2002) and program processes to evaluate implementation fidelity (Dusenbury et al. 2003).

[7]Pedhazur and Schmelkin (1991) provide an elaborate description of measurement validity, reliability, and approaches to testing these characteristics, which could be consulted to gain a foundational understanding of measurement theory and practice.

conclusion may be that the program had an effect in reducing drug use in established users, but did not have an impact on the normal developmental trend of experimentation for nonusers.

Finally, because of the floor effect of most drug use outcomes (most students report no use), evaluators tend to collapse the responses to these variables into a dichotomous measure of either "no use" or "used once or more." This approach results in the loss of even more information and limits the evaluator in the type of change question he/she can answer to whether there is an increase in the initiation of use or an overall decrease in the prevalence of use. Lost is the ability to detect changes in the average frequency of use. It is therefore critical to select measures that represent the constructs of interest accurately and reliably, and that will be sensitive to any changes produced by the program. Additionally, other methods for increasing the power to detect a program effect include the regression of post-test scores on pre-test scores, increasing the sample size or the alpha level for hypothesis testing (Mohr 1995), measuring and controlling for covariates, using multiple measures of the construct, and employing analytical techniques such as latent variable modeling to partition the error variance in the outcome variable (Shadish et al. 2002).

The evaluator should recognize that measurement issues, as well as the design of the study, are intricately tied to the selection of data analysis techniques utilized to test hypotheses regarding the program's effect on mediators and outcomes. For example, a causal design leads the evaluator to test hypotheses about differences in groups on the targeted mediator or outcome, as well as change over time in those constructs. Experiments require fewer constructs to be measured than quasi-experimental designs because the design (randomization to an intervention or control group) controls for confounding influences that pose threats to internal validity. In addition to how many constructs are measured and the research question or hypothesis being tested, the level of measurement of the variables being analyzed determine the specific analytical technique utilized to address the research question or hypothesis. Additionally, the number of data collection points, the length of time between those measurement points, the unit of program assignment and sampling, the unit of analysis, and the theoretical relationship between mediators and outcomes all present measurement issues that should be considered and planned while explicating the program theory and planning the evaluation (Collins and Flaherty 2006; Collins 1994).

20.6 Conducting/Monitoring Evaluation Procedures

Monitoring and documenting program inputs and activities are critical components of a quality evaluation; again, the logic model is a useful tool for determining what activities are undertaken, and therefore should be monitored to achieve the program goals. The evaluator's role is to provide leadership in organizing and observing those activities and documenting, through observation and data collection, the extent to which the necessary activities are completed. Prior to the evaluation, the evaluator should create detailed documentation of who is responsible for each activity, including logistical arrangements for the training of key personnel, construction of data collection instruments and procedures, data storage and analysis, and the production and dissemination of findings.

As noted earlier, program evaluation can be conducted externally, internally, or through some hybrid approach. Traditionally, external evaluations that are conducted by an independent third party have been viewed as more objective and valid, while evaluations conducted by internal personnel have often been viewed as potentially suspicious and lacking methodologically in that evaluators can be pressured (biased) to show positive program outcomes (Torres 1991). Hybrid approaches that utilize both internal and external components are seen as compromises that seek to increase the scientific rigor of an evaluation while containing costs and addressing other issues. For example, an organization may hire a consultant to design an evaluation and analyze and report evaluation findings, but utilize internal personnel

to manage the evaluation, including the collection of data and/or other aspects of the evaluation. Conley-Tyler (2005) offers guidelines to assist in choosing internal or external evaluators. Regardless of the approach taken, objectivity and sound evaluation methodology are vital.

Few, if any, public health programs can be evaluated within the confines of a laboratory, where strict control is possible to eliminate confounding variables that may impact the results of the evaluation. Even randomized controlled trials (RCT), considered the gold standard of experimental design, are often not possible when assessing the effectiveness of public health programs. Consequently, evaluators must diligently manage the conduct of field evaluations in an effort to minimize the impact of influences that may reduce confidence in results, including issues of internal validity, properly trained personnel, valid and reliable measures, and well-developed data collection protocols (and strict adherence to those protocols) all help to ensure that an evaluation is conducted properly and objectively.

Prior to an evaluation, evaluators should create detailed documentation of who is responsible for each activity, including logistical arrangements for training of key personnel, construction of data collection instruments and procedures, data storage and analysis, and the production and dissemination of findings. Tools such as a Task Development Timeline (TDTL) or Gantt chart (Gantt 1974) are useful in the management of an evaluation, in terms of ensuring that tasks are delegated and completed on time, and that goals and objectives are met. If not properly implemented and managed, the best-designed and most rigorous evaluations can become problematic and thus reduce the level of confidence associated with the results.

Whether internal or external personnel are utilized to conduct an evaluation, those involved should be properly trained. External evaluators are ideally professionals with substantial and relevant experience in conducting evaluations. These professionals often include, but are far from limited to, academics affiliated with a university. As such, background and experience should be sufficient. However, organizations seeking external evaluators should carefully review candidate experience and fit with regard to the agency and the type of program to be evaluated. Personnel internal to an agency will likely have varying degrees of experience and knowledge regarding program evaluation. At a minimum, internal personnel should have a general understanding of program evaluation, including the specific purposes (demonstrate efficacy, improve program) of the evaluation being conducted. The importance of objectivity should be stressed. Furthermore, evaluation protocols, including protocols for data collection and entry, should be developed and strictly adhered to. Supervision of personnel should be ongoing and include regular audits to ensure that protocols are being properly followed. Additional training sessions may be necessary over time, especially if new or different personnel are brought on to assist with the evaluation.

Collecting and processing data for an evaluation must be done carefully and accurately. A program may actually be very effective in changing behavior, but may not appear to be effective due to flawed data collection and/or processing. Instruments used to collect program evaluation data, such as measures of risk and protective factors or substance use, vary in terms of their difficulty to administer; training of data collection staff should be appropriate for the type of data collection utilized to insure the validity of the data. For example, some measures are self-administered and only require the respondent to check or circle responses. Other measures must be administered by trained professionals via face-to-face interview. In choosing or developing a measure, several considerations should be made: are the questions relevant to the target population? Is the measure culturally appropriate? Can the target population read and understand the questions? These, as well as many other considerations may have a significant impact on the quality and validity of the data collected.

Data processors and analysts should also be trained to work with the data as scoring of measures range from simple summation of individual responses to more complex computations involving statistical transformations. This is especially

true with qualitative and mixed-method designs, as both are common in evaluation. For instance, proficient and defendable procedures for analyzing qualitative data vary considerably, based on a range of analytic traditions for qualitative research, from narrative analysis to "critical realist" epistemology (Thomas 2006). Analytic strategies or approaches further vary by whether an inductive or deductive approach is selected, as both are conventional in qualitative evaluation.[8] Integrative strategies for mixed-method analysis can also be challenging. Choices range from transforming one data type (usually qualitative) to allow for statistical or thematic analysis of both together to actual data merging, which involves more sophisticated use of joint data (via coding and other iterative methodologies) to create new or consolidated variables/data sets (Caracelli and Greene 1993). On the quantitative side, another important consideration is how to handle missing data in field research. While beyond the scope of this chapter, it is important to note here that methodologies and software exist to perform missing data analyses and diagnostics, including strategies for reducing bias effects associated with methods of handling missing data and frameworks to deal with nonrandom missing data (Graham 2009; Raymond 1986). Overall, while not exhaustive, this discussion provides some insight on the considerations that must be given to data reduction and analytic strategies when conducting evaluations of substance abuse prevention and treatment programs.

To conduct an evaluation but not report or disseminate findings would largely undermine the intended purpose of program evaluation. According to the Centers for Disease Control and Prevention (CDC), "…evaluation results can be used to demonstrate effectiveness of your program, identify ways to improve your program, modify program planning, demonstrate accountability, and justify funding" (DHHS 2011). Evaluation results can also demonstrate that resources are being spent appropriately, show changes over time, focus attention on new or important issues, promote the program, assist in obtaining or increasing funding, and provide direction to program staff (DHHS 2011). Fundamentally, a program evaluation report should be able to answer the research questions originally identified at the onset of the evaluation (did the program work?). When reporting and disseminating the results of an evaluation, one should consider the intended recipients of the information. For instance, will the report be read by academics and other researchers, or by laypersons (agency administrators, stakeholders)? A report written for agency administrators will likely need to include descriptive information written in plain language. Reports should follow the general format of a peer-reviewed scientific paper (abstract/executive summary, introduction, methodology, results, discussion, and consideration of limitations) and include graphical representations of results when appropriate. Prior to the dissemination of findings, it is often wise to solicit feedback and comment from stakeholders regarding the content and appropriateness of the report, especially with regard to findings that are negative or contrary to expectations.

20.7 Conclusion

This chapter provides a brief overview of processes and considerations important in the evaluation of substance abuse prevention and treatment programs. The chapter Appendix provides an example of an actual evaluation of the Adolescent Substance Abuse Prevention Study (ASAPS) as a means to further illustrate the processes described in this chapter.

Appendix: Application Example—The Adolescent Substance Abuse Prevention Study (ASAPS)

The evaluation of the Adolescent Substance Abuse Prevention Study (ASAPS) was conducted with the goal of assessing the

[8]Choice of approach has implications not only for basic analytic strategy, but also data coding procedures, methods for data verification and reporting (Patton 1999; Thomas 2006).

implementation and effectiveness of a two-component universal, school-based substance abuse prevention curriculum delivered by police officers who had previously been trained, and were currently teaching the Drug Abuse Resistance Education (D.A.R.E.) program. The research team for the evaluation were located at The University of Akron Institute for Health and Social Policy in Ohio, and stakeholders who collaborated with the team included D.A.R.E. America organization leadership, trainers, officers, and students, substance abuse prevention researchers, research methodologists, and statisticians from across the United States, curriculum specialists and teachers, and the Robert Wood Johnson Foundation (RWJF).

Evaluation planning begun in 1999 when RWJF funded a project to revise and evaluate the D.A.R.E. curricula currently being implemented in schools. At that time, the principal investigator contacted and invited stakeholders, researchers, and educators to participate in two planning groups: (1) a curriculum workgroup, and (2) a research design workgroup. Both groups worked concurrently on planning curriculum revisions, as well as the study's evaluation design. This process took nearly 2 years to complete and included pilot studies, capacity building for the administration of the study, and school and police department recruitment in six large cities across the United States. The new curricula, Take Charge of Your Life (TCYL), was implemented in 120 middle schools in six sites across the continental U.S. and data collection for the outcome evaluation study began in the fall of 2001. The 9th grade component was delivered to the same cohort of students in the 2003/2004 academic year. Students were followed annually for data collection until they were in the 11th grade (2005/2006).

Defining Program Goals and Processes of the ASAPS

The overarching goal of the new curriculum was to delay substance use initiation or reduce current levels of use in an ethnically and socially diverse U.S. middle and high school populations. Curriculum materials were developed by prevention experts using criteria for effective prevention programming derived from existing meta-analyses and reviews of the literature. The curriculum workgroup used these final recommendations to develop the new middle and high school curricula, using a problem driven format based on authentic dilemmas and issues faced by teens as they are pressured or tempted to experiment with or use tobacco, alcohol, illegal drugs, or inhalants. The primary instructional strategy used student-to-student engagement through instructor guided in-depth discussions, role-playing of skills and concepts, and small group problem-solving.

The chosen framework enables students to actively utilize the intended ideas and skills as they develop their own understandings and capacities to be in control of situations where they are pressured to use tobacco, alcohol, and drugs. In order to attain these objectives, the curriculum focused on the following specific targets/constructs for change (immediate and intermediate outcomes) at the student level:

- Consequences of substance use: understand the nature of and risks (personal, physical, social, legal) associated with the use of alcohol, drugs, tobacco, and inhalants.
- Beliefs and attitudes toward substance use: examine and understand their own beliefs and attitudes related to alcohol, drug, tobacco, and inhalant use.
- Normative beliefs regarding peer use of substances: correct misperceptions about the rates of substance use by same-age adolescents.
- Decision-making skills: make positive quality of life decisions.
- Communication skills: communicate clearly and interact positively in social and interpersonal situations.
- Resistance/refusal skills: develop and use resistance skills.[9]

[9]Skills are available on the Robert Wood Johnson Foundation Website: http://www.rwjf.org/en/library/research/2009/06/the-adolescent-substance-abuse-prevention-study.html.

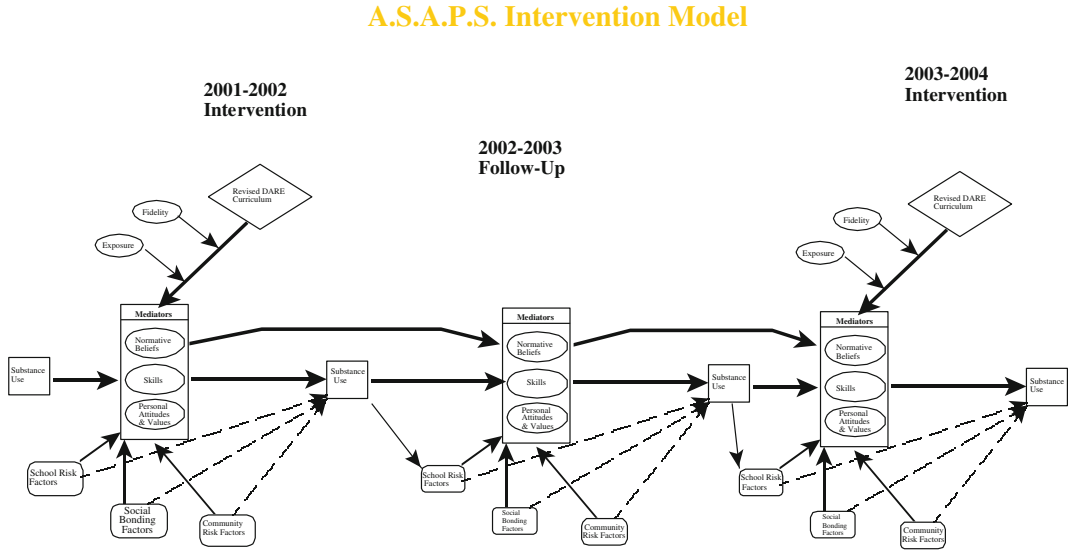

Fig. 20.2 Program theory for adolescent substance abuse prevention program

The final curricula was composed of ten, 40 min lessons administered in the 7th grade, and seven, 40 min sessions delivered to the same students (cohort) when they were in the 9th grade. These targeted constructs were understood to operate within a larger, ecological set of influences ranging from family to community levels. The result of research and development of the two-part curricula was the theoretical model which guided both the intervention as a whole, and the research design for evaluating the program. This theoretical model is depicted in Fig. 20.2.

In this model, substance use is seen as a direct outcome of the students' knowledge and beliefs about the nature and effects of various dangerous substances, their attitudes toward the use and consequences of abuse of the various substances, and the level of personal skills which may allow them to effectively resist pressure to participate in the use of any illicit substances. Thus, the intervention targeted for change the student's normative beliefs, skills (immediate outcomes), personal attitudes and values (intermediate outcomes) which in turn were expected to reduce the student's intentions to use substances (intermediate outcome) and substance use (ultimate outcome). As can be seen in Fig. 20.2, these constructs are conceptualized to mediate the effects of the intervention on substance use. For this reason, the mediating constructs/variables are treated as proximal/immediate or intermediate outcomes of the intervention, while substance use is viewed as the distal/ultimate outcome.[10] In order to create a more comprehensive framework for a substance prevention program, the model also included constructs thought to directly influence the student's normative beliefs, communication skills, and personal attitudes and values and indirectly influence the student's substance use. These school, community and student risk indicators were included and are represented in Fig. 20.2 as social bonding, school, and community risk factors (National Institute on Drug Abuse 1999).

Formative Evaluation

The formative evaluation of the new, 7th grade curriculum began with separate focus groups of 7th grade students, their parents, middle school

[10]The model also included the constructs of implementation fidelity and exposure. Both of these constructs were thought to act as moderators of the effect of the intervention on the student's normative beliefs, skills, and personal attitudes.

teachers, and D.A.R.E. police officers to elicit feedback on the curriculum goals, content, and materials. Students were asked to provide input on ways to revise the problematic situations presented, as well as for the inclusion of any other problems they felt were relevant to same-age adolescents. In addition, students were queried about their opinions regarding having police officers teaching in the classroom. Parents' attitudes about drug abuse, prevention programming, and having police officers in the classroom were also explored. The same process was followed with local police officers who had experience with the D.A.R.E. program. Feedback on the ability/motivation of officers to implement this type of lesson and their ability to involve students in active discussion of the problems presented were solicited. The officers were also asked to evaluate how realistic each problem was and the accompanying activities and to provide input/insight into problematic situations they have been exposed to in the D.A.R.E. classroom. These problems were incorporated into the lesson plans. Finally, middle school teachers were recruited to review the lesson plans and instructor training manual.

The first trial delivery of the middle school curriculum (pilot study I) was tested in one urban middle school. All 7th grade health or social studies classes received the middle school curriculum (n = 153). All data collection instruments were tested in the feasibility study including measures of outcome and mediating constructs, fidelity of implementation, and exposure. Two local D.A.R.E. officers participated in a two day training on the curriculum theory, content, intended delivery procedures, and background information; each was responsible for half of the 7th grade classes. In order to gather as much information as possible, each class was observed by trained raters from the research institute. Officers were observed by one or more members of the curriculum workgroup during delivery of each lesson. Feedback was exchanged on the lesson and revisions were made to the lessons based on these sessions and the feedback from the focus groups.

To test the specific objectives of the study, a number of study instruments were developed, including instructor observation sheets, short, lesson specific, evaluation surveys for instructors and students, a student survey with measures of the targeted outcomes and mediators, risk and protective factors, and demographic information. These instruments relied heavily on existing measurement instruments, especially those measuring students' substance use as well as normative beliefs, decision, communication and refusal skills, and personal attitudes and values related to substance use. Based on an extensive review of the literature, it was evident that in addition to using many existing measures, new or adapted measures of many of the study constructs would be necessary. These newly developed measures were pre-tested in the trial delivery of the middle school curriculum by the institute's research staff.

A second pilot study was conducted in nine middle schools using a non-randomized, no-comparison group design; a convenience sample of nine schools, including urban, suburban, and rural public schools and one private (Catholic) school. Data on mediators and outcomes were collected from 462 students. The results of the second pilot study were promising with significant changes in the expected direction for the immediate outcomes of normative beliefs, refusal skills, and the perceptions of harm for ATOD use. Instructors implemented the program with high fidelity, and feedback was positive toward the feasibility of implementing the program on a larger scale. The positive results of the two pilot studies indicated the curriculum and officer training were ready for the larger efficacy study.

Outcome Evaluation

The research workgroup implemented a randomized, longitudinal control trial, to evaluate the intervention. School districts (middle and high school clusters) were randomly assigned to either a treatment (intervention schools that would receive the new TCYL curricula) or a

control condition. However, ethical concerns prevented the study group from requiring the control condition schools from refraining from implementing any prevention programming; consequently, the control condition became the condition of 'prevention programming as usual.'

Sampling

High schools and all their feeder middle schools were selected as the clusters of interest, with the school district becoming the unit of sampling. A sample size of approximately 40 school districts per condition was expected to provide statistical power of 0.80, assuming an alpha level of 0.05 and a two-tailed test of hypothesis. To insure diversity in ethnicity and geographic location, six large metropolitan areas were selected for recruitment, with the 'core' central city school district identified as the center point for a 50 mile sampling radius (with the goal of including urban, suburban, and rural school districts) for each region of the United States. These regions were centered in the following metropolitan areas: (1) New York City, NY/Newark, NJ,[11] (2) Detroit, MI, (3) St. Louis, MO, (4) New Orleans, LA, (5) Houston, TX, and (6) Los Angeles, CA. A sampling frame was developed for the area surrounding each 'core' district, with the inclusion criteria of public school districts with middle schools that housed 7th and 8th graders and four-year high schools. To ensure diversity across schools, the six sampling frames were stratified by risk status, which was calculated using a function of the proportion of students eligible for free lunch (an indicator of SES/poverty) and minority enrollment in the school.

[11]New York City was already signed on as the core school district in NY, but the bombing of the twin towers (911) occurred while the other NY city schools were being recruiting resulting in a loss of potential and current schools in the study. To increase the available sampling frame, the potential sampling area was expanded to include Newark and surrounding schools in NJ.

Organizational Structure of the Study

Given that this was a large evaluation study that spanned six metropolitan regions of the United States, a well-planned administrative structure was a critical component for a successful implementation of the study plan. Two principal investigators oversaw all the study activities, a senior researcher oversaw the sampling and assignment of districts to intervention condition, another senior researcher oversaw curriculum implementation and monitoring, and a senior researcher oversaw data analysis. A project manager coordinated the recruitment of school districts and police departments and training of police officer/instructors, and a data manager oversaw data collection and processing. In addition, regional coordinators were hired to monitor on-site recruitment, communication, data collection, and retention of schools and subjects. They also supervised full-time site coordinators for each of the six regions who oversaw part- and/or full-time data collection personnel and acted as liaisons between the schools, police, and community, while supervising curricula delivery and data collection procedures.

Recruitment and Retention of Schools

The local recruiting process was then directed by the site coordinator in each region. An explanatory brochure and a cover letter were sent to the selected school district superintendents requesting their participation in the study. They were informed that their agreement to participate in the study would not guarantee that their schools would receive the program and that if they are included in the control group, they would not be able to deliver this program for two years to allow the 7th grade cohort to transition to their high school. The mailings were followed with a telephone call from the Project Manager at the Institute; this correspondence provided information to the coordinator as to the gatekeepers and

appropriate person(s) with which to continue recruitment of the district schools. The actual recruitment activities for each school district varied; however, the process generally included site visits by senior research staff members to explain the study to district personnel, principals, teachers, and parents/PTA associations, along with visits to the local police/sheriff's department to recruit local officers to train for instructors for the TCYL curricula.

School districts were required to agree to be assigned to either the control or intervention (treatment) condition to participate in the study. Once district personnel agree to participate, similar materials, a letter of support from the district superintendent, and phone calls were made to the principals of the middle schools and high schools to elicit their cooperation. In recognition of the costs associated with participation with the study, each school was offered $500 per year of participation in the study and resource materials relative to substance abuse prevention as incentives to participate. They were also promised copies of the final study report.

After all agreement letters were signed, the district was randomly assigned to either the intervention/treatment or control condition. The core districts were randomly assigned first (to achieve a balance of control and intervention inner-city districts) and the schools surrounding the 'core' district were assigned randomly without any attempt to achieve a balance of control and intervention districts (although the sample was fairly balanced within each region).

Retention of schools was one of the primary roles of the site coordinators. In addition to the incentives, each regional site coordinator convened a 'Community Advisory Group,' consisting of local community leaders representing the schools, business community, local political groups, social service agencies, community coalitions, and/or law enforcement. The Advisory Group's role included supporting the project within the community, assisting in the interpretation of study findings, and serving as a support network to establish the program in the middle and high schools after the completion of the research project in both the experimental and control communities.

Recruitment, Retention, and Tracking of Students

Parents and students in both the intervention and control schools were required to sign a consent form to participate in the survey administration (all intervention students received the curriculum, regardless of survey consent status). Parents and students were also asked to provide information for persons who would know the whereabouts of the student for follow-up purposes. Parents and students were reminded about the need for the signed consent forms and incentives; to increase signed consent forms, incentives were offered to students in terms of classroom-wide activities, such as pizza parties. The consent forms were available in both English and Spanish.

One week before the first intervention class (7th and 9th grades), a baseline survey was administered to students who had provided appropriate consent. To ensure anonymity, student was assigned a unique identifier. This key file was then stored separately from any student data for use in matching coded student surveys to the data file at each data collection point. Once the students completed the survey, they placed the coded survey forms in the blank envelope, sealed the envelope, and deposited it in a slit on a sealed box that was located at the front of the classroom; this procedure was used at each data collection point.

Site coordinator and team members were responsible for student tracking. Students who were not present at a data collection session but were still enrolled in school were approached individually by the site coordinator and asked to complete the forms in an empty classroom or other private place to ensure confidentiality. Students who had left the district were lost to

follow-up. The decision not to follow-up students who left the school district was made after an attrition pilot study was conducted by the research group and the findings indicated that the cost for locating these students was prohibitive (Stephens et al. 2007).

Data Collection and Measures

The data collected can be summarized by five general construct categories: (1) substance use outcomes, (2) mediating outcomes, (3) risk factors, (4) moderators of the intervention, and (5) indicators of implementation fidelity. The outcomes and individual risk factors were measured as part of the main student survey; other variables required data collection from additional and sometimes multiple sources, such as the officer/instructors, curriculum trainers, site coordinators and ethnographer's interviews with students, and key community and school informants. As there were many sources of data collected for this evaluation, this example will focus only on data collected from student surveys (outcome evaluation) and implementation fidelity observations (process evaluation).

Baseline student surveys were administered immediately prior to program delivery in 7th grade, and a post-test was administered approximately 90 days after the completion of the ten session curriculum. A third post-test was administered in the 8th grade, and pre- and post-tests were administered before and after the implementation of the 9th grade curriculum (seven sessions). Post-tests in the 10th and 11th grade were administered approximately one and two years after the 9th grade post-test. Neither the officer/instructor nor the classroom teacher was present during data collection to assure students that their responses on the survey would be confidential. To maintain generalizability with national data systems and other prevention program evaluations, survey measures were adapted from ongoing studies, such as Monitoring the Future Study (Johnston et al. 2013), Center for Substance Abuse's Core Measures (1999), and prior studies of D.A.R.E wherever possible. Any changes to the original measures were tested for validity and reliability in the feasibility study.

Since the curriculum is specifically aimed at reducing or preventing the use of alcohol, tobacco, marijuana, and inhalants, this construct was measured using the self-report of students' use of tobacco, alcohol, marijuana, and inhalants on a paper-pencil, confidential student survey at seven points in time. The age of first use as well as pattern of use (lifetime, last 12 months, and last 30 days) and amount of use where appropriate (binge drinking) were measured using questions taken from the Monitoring the Future survey. To measure age at first use of substances, students were asked, for example, "How old were you the first time you had a full drink of an alcoholic beverage?" with ordinal response options ranging from "never" to "15 or 16." Substance use was operationalized with questions such as, "How many TIMES (if any) have you had alcoholic beverages to drink (more than just a few sips)…." "during the last 12 month" or "during the last 30 days." Seven response options ranged from "never" (coded as zero), to "40 or more times" (coded as 6).

With regard to mediating variables, intentions to use substances was measured with drug-specific questions, asking how likely the student was to try alcohol, tobacco, or marijuana in the next 12 months. Responses ranged from 1 (definitely will) to 5 (definitely will not). Attitudes toward the use of tobacco, alcohol and marijuana were measured by two items for each drug. Students selected a response to complete the stem, "I think it is okay for students my age to…" Response items included, "smoke cigarettes once in a while," "drink alcohol almost every weekend," and "smoke marijuana once in a while." Responses ranged from 1 (agree) to 5 (disagree). Normative beliefs were measured by three questions that asked students how many 10th graders the student believed used tobacco, alcohol, or marijuana in the last 30 days. Response categories ranged from 1 (more than 75%) to 5 (10% or less). Perceptions of harmful consequences resulting from substance use were measured by three items that asked students how much they thought the use of a particular

substance (alcohol, tobacco, or marijuana) affects how the brain works. The response categories ranged from 1(none) to 5 (a lot).[12]

Refusal skills were assessed by responses given to three hypothetical scenarios involving the opportunity to use tobacco, alcohol, or marijuana. Students were asked to read the scenario in which a substance was offered by a peer. Each student selected the best refusal response to that offers from a list of possible responses given the person being offered does not want to use the substance. Responses were weighted according to the level of assertiveness demonstrated. For example, a response of "no, maybe later" was assigned a lower score than a response of "no thanks, I don't want to smoke." Scores ranged from 0 (least appropriate response) to 2 (best response chosen). The student survey also included risk and protective factors and demographic characteristics including self-reported age, sex, race/ethnicity, and family composition.

Implementation Fidelity

Fidelity of implementation was measured by the amount of intervention material actually covered in each session, the number of intervention sessions completed, the amount of role-playing, demonstration, and discussion that occurred during a sample of two sessions, the number of times the instructor reinvented or altered the material or delivery, and overall quality of delivery. These constructs fit with the recommendations of Baranowski and Stables (2000) that implementation and reach are critical components of process evaluation. These data came from four sources: (1) independent classroom observation checklists completed by trained site staff, (2) post-instruction self-report surveys by officer/instructors, (3) post-instruction self-report surveys by students, and (4) attendance records for each student in each of the intervention sessions.

Data Processing and Analysis

Data collection was overseen by the site coordinators and their team members. Student surveys were scored/scanned electronically and sent to a master data base maintained by the Institute to assure the highest possible quality. Data were assessed for completeness, that responses were within range of the values for each variable, and that data were internally consistent. Particular attention was paid to assuring that the longitudinal data on the cohort sample could be linked across all waves of data. As soon as data were received, linkages were established and Site Coordinators were contacted to remedy any discrepancies. Data from observation forms and implementation fidelity surveys were also scanned and cleaned at the Institute and stored electronically.

Data Analysis

The unit of sampling was the school district, but the curricula were delivered in classrooms within those schools, and the surveys were administered to students. Hence, students were nested within the classroom, school, and district. This complex sampling design presented the potential for biased estimates for the standard errors of any hypothesis tests. Therefore, all analyses were conducting using statistical procedures which adjusted the standard errors.[13] Logit models were

[12]Survey items also measured decision-making and communication skills. Decision-making skills was measured using items drawn from a scale score developed by Goldstein and McGinnis (1997). Scores reflected the student's rating (on a scale from 1 = never to 5 = always) of statements such as, "Before making a decision, I think about all the things that may happen as a result of that decision." (alpha > 0.70). A composite measure of communication skills, taken from the Social Orientation Scale (Cegala 1981). Items measured the perceptions interpersonal communication confidence and competency with statements such as, "I feel confident of what to say and do during conversations." Responses were on a Likert scale of 1 (disagree) to 5 (agree) (alpha > 0.70).

[13]For further reading on the problem of nested/complex sampling and standard errors please consult Raudenbush and Bryk (2002).

utilized for binomial outcomes, in addition to weighted least squares, while maximum likelihood models were used in path modeling and structural equation models with outcomes at the ordinal and interval levels.

Evaluation Study Findings

Process Evaluation (Implementation Fidelity)

The process evaluation reported here focuses on the implementation fidelity of the officer/instructors delivering the new curriculum in the 7th and 9th grades and the exposure of students to the curriculum (as measured by lessons attended). Classroom observations conducted in the 7th grade and 9th grade all showed high implementation fidelity. Coverage of the content of the lessons in 7th grade ranged from 34 to 100% with a median of 72 and 81% content coverage for each of the two lessons observed. For the 9th grade, the proportion of content covered in each of the two lessons observed ranged from 12 to 100% with a median proportion of 70 and 78% for each of the observed lessons. Instructors' use of the appropriate activity in their instruction was somewhat lower, with a median proportion of 63 and 44% delivering the two observed lessons with the appropriate instructional activity in the 7th grade and 50 and 60% using the appropriate instructional strategy during the observed lessons in the 9th grade (Sloboda et al. 2009a, b). While these numbers may seem low, they were relatively high compared to other studies reported in the literature (Ennett et al. 2011; Hallfors and Godette 2002; Ennett et al. 2003).

Attendance records indicated there was adequate exposure to the curricula with 69% of 7th grade intervention students attending all ten lessons and another 27% of students attending eight or more of the ten lessons. In the 9th grade, 44% of the intervention students attended 100% of the seven lessons and 17% of the intervention students attended at least five (71%) of the seven lessons.

Outcome Evaluation

The outcome evaluation was completed when the cohort of students were 11th graders. A detailed reporting of the data analytical procedures and finding is reported in Sloboda et al. (2009a, b). Forty-two high schools and 59 middle schools were assigned to the intervention (TCYL curriculum intervention) and 41 high schools with their 63 feeder middle schools were assigned to the control condition. A total of 19,529 students completed consent forms prior to the administration of the 7th pre-test (intervention group $n = 11{,}314$; control group $n = 8215$). Baseline surveys were completed by 10,028 intervention students and 7302 control students. During the course of the study, three high schools were lost to follow-up—two were destroyed in Hurricane Katrina and one opted out of the study. Baseline data were used to check the randomization process between the intervention and control schools; demographic characteristics and substance use outcomes showed no significant differences between the groups, thus confirming the randomization procedure was successful. Attrition analyses at the 11th grade post-test showed that overall, older students, female students, non-White students, and students who reported the use of alcohol, tobacco, or marijuana were more likely to drop out of the study. The only source of differential attrition was that students who identified as "other-race" were more likely to drop out of the control group than the intervention group (Sloboda et al. 2009a, b; Teasdale et al. 2009).

The primary goal of the intervention was to reduce or delay the onset of substance use among the cohort of students who received the TCYL curricula. The program theory proposed that the curricula would change the targets of normative beliefs, consequences of substance use, attitudes towards substance use, refusal, decision-making, and communications skills (immediate outcomes/mediators). Changes in these constructs would result in students' intentions to avoid substance use (intermediate outcome) and changes in intentions would result in lower use of tobacco, alcohol, and marijuana (ultimate

outcome). The findings did not support the hypothesis that the curriculum would have an impact on these outcomes. In fact, although there had been intervention effects on normative beliefs and refusal skills when students were in middle school, these effects were no longer significant in the 11th grade. There was no intervention effect on past 12-month outcomes of alcohol use, getting drunk on alcohol, or marijuana use. Surprisingly, alcohol use and getting drunk on alcohol in the past 30 days showed a significant effect in the opposite direction as expected, as did the 30-day use of tobacco and marijuana and the self-reported past two weeks binge drinking outcome. That is, the intervention group reported, on average, using tobacco, alcohol, and marijuana at higher rates than the control group in the 30 days prior to the 11th grade survey administration. The risk ratios for these effects ranged from 1.09 for alcohol use to 1.21 for cigarette smoking. In order to further understand these puzzling findings, subgroup analyses was utilized to test the moderating effects of prior substance use (as reported at baseline), sex, and race/ethnicity were conducted.

The analyses by substance use status, sex, and race/ethnicity did provide the evaluators with some insight for whom the program was and was not successful in reducing substance use. Males in the intervention group appeared to be driving the higher rates of alcohol use, getting drunk and bingeing on alcohol, while females in the intervention group were more likely (than their control group counterparts) to binge drink and smoke cigarettes. Non-White students in the intervention group had higher rates of cigarette use than non-White students in the control group, and White students in the intervention group had significantly higher alcohol outcomes than their White counterparts in the control group. The most surprising findings were those found when the nonusers at baseline were compared to substance users at baseline. Students who reported no use of alcohol at baseline in the intervention group were more likely than nonusers of alcohol in the control group to report binge drinking in the past two weeks, and alcohol use and getting drunk on alcohol in the past 30 days. Similarly, non-smokers in the intervention group were more likely than nonsmokers in the control group to report smoking cigarettes at the 11th grade post-test (Sloboda et al. 2009a, b).

A single finding regarding marijuana use was found in the expected direction for the intervention group. Students in the intervention group who reported using marijuana at baseline had significantly lower rates of use in the 11th grade than did non-marijuana users in the control group. In summary, the intervention appeared to work only in early marijuana users to reduce marijuana use. The intervention also appeared to have the effect of increasing smoking in females and non-Whites students, increase drinking in Whites and males, increasing problem drinking in females, and increasing smoking, alcohol use, and problem drinking in students who were nonusers in the 7th grade (Sloboda et al. 2009a, b).

What Went Wrong?

Two analytical questions were explored in an effort to provide insight into the study results. First, the basic program model of immediate, intermediate, and ultimate program targets was assessed to determine if the curriculum was targeting constructs that, if changed, would change the ultimate outcome of substance use. Path modeling was used to examine the relationships among normative beliefs, attitudes toward use, perceptions of consequences of substance use and refusal, communication, and decision-making skills on the intentions to use and actual use of cigarettes, alcohol, and marijuana. The findings proved interesting in that the effects (direct and indirect) of these mediators on each of the outcomes differed slightly; for example, refusal skills worked to reduce intentions to use cigarettes and marijuana by helping students to utilize decision-making skills to form intentions, but impacted alcohol use directly and possibly only for students who already had formed negative attitudes toward alcohol use. However, with the exception of communication skills (which appeared to increase intentions to use alcohol, and

had no effect on cigarette or marijuana use), all program targets had a significant indirect or direct effect on each of the substances, with normative beliefs having the largest total (indirect + direct) effect on each substance. The effect of each of the program targets on substance use was small, and therefore the program would have to produce large changes to actually have an effect on the substance use outcomes (Stephens et al. 2009).

In fact, as noted above, while the program showed effects early on (in middle school) on normative beliefs, perceptions of consequences and refusal skills, by the 11th grade, these effects had disappeared and none of the targeted mediators showed differences between the control and intervention group. The second set of analyses used path modeling to examine the relationship of the program model constructs for baseline users and nonusers; these analyses were done to explore the single positive effect of the program on baseline marijuana users. The results of these models showed several significant effects of the treatment variable on the targeted mediators which were not seen in the intervention group overall. For nonusers, the intervention had a significant effect only on two marijuana specific mediators. Nonusers in the intervention group were significantly higher on their 9th grade marijuana refusal skills and marijuana specific normative beliefs than nonusers in the control group. For the baseline users, intervention effects were shown on mediators for each of the three substances. Baseline cigarette users in the intervention group were significantly higher than baseline users in the control group on the perceptions of harm for cigarette use. Baseline alcohol users in the intervention group were significantly higher than baseline users in the control group on normative beliefs about alcohol use and the perceptions of harm for using alcohol. Finally, baseline marijuana users in the intervention group were higher on their intentions not to use marijuana, marijuana refusal skills, and normative beliefs about marijuana use.

The TCYL intervention appeared to have no effect on the cigarette specific mediators in the nonuser group, but the program did appear to have a significant direct effect on cigarette use in the direction of higher use for TCYL baseline nonusers. There was, however, a significant program effect for the baseline user group on cigarette specific perceptions of consequences. The findings were similar for alcohol use, with the addition of a significant program effect on normative beliefs surrounding the use of alcohol for students who were baseline users of alcohol. The results for the model of marijuana use are similar for both baseline users and nonusers. The main effect of the program on marijuana use and the intentions not to use marijuana for baseline users became nonsignificant when the mediators are included in the same model, indicating full mediation of the program through marijuana specific refusal skills and normative beliefs, both of which showed positive program effects for baseline nonusers and users. Neither baseline users nor baseline nonusers showed any program effects on the global measures of communication or decision-making skills (Teasdale et al. 2009).

These findings partially explained why the intervention had beneficial programmatic impacts on marijuana use for baseline users. That is, the TCYL intervention reduced beliefs about the normative nature of marijuana use and increased refusal skills, compared to control students. In contrast, no explanation was found for why the program had the negative impacts for nonusers. It is interesting to note that students in the TCYL program did not report increased intentions to use alcohol, tobacco, and marijuana relative to control students. Based on the program theory, any impacts of the program should have worked through the targeted mediators (normative beliefs, consequences, and skills) and behavioral intentions. This was not the case, leaving the question of what did the intervention do to create the negative outcomes for baseline nonusers? If it was not the proposed theoretical mediators that influenced these outcomes, what components of the intervention impacted substance use? The evaluation group and other researcher continue to explore these findings to

determine what went wrong and whether or not the intervention can be improved to change the targeted constructs.

Lessons Learned from the ASAPS

There are many lessons to be learned from the ASAPS, but this chapter focuses on those regarding the evaluation process. First, the ASAPS as an evaluation undertaking was a success in that the evaluation team collaborated with a multitude of stakeholders to come to agreement on program goals and processes, as well as how to evaluate those processes and outcomes. The shared responsibility and input of these stakeholders provided the framework for a comprehensive evaluation, the results of which continue to be utilized by substance abuse prevention specialists, educators, and public health professionals.

Second, this study illustrates the importance of incorporating program theory into the evaluation process and measuring and analyzing constructs and processes that compose that theory. While the evaluators expected the program goal of reduced substance use to be met, they were, with only one exception, not achieved. To understand this contradictory finding, they had the measures of the program theory and implementation procedures to examine to determine why the intervention did not work as anticipated. They found that while the program did impact some of the proposed mediators of substance use, the changes were not consistent or large enough to have an effect on substance use across a subpopulation of students. The program was implemented with fidelity, so this did not appear to be the weakness. Perhaps the content of the curricula was not powerful enough to change the targeted constructs? The lesson learned is that the program should be revised to strengthen its impact on these constructs before being implemented in middle and high schools.

Finally, the importance of utilizing program evaluation findings is also illustrated in this study. As a result of these analyses, the evaluators came to two important conclusions: (1) the TCYL curricula was not a universal curricula, and would be more appropriate for high risk students, and (2) it may suggest to focus on targeted subpopulations of students (users OR nonusers) rather than using "universal" programs that are delivered to a diverse population of students. These conclusions were taken seriously by the decision-maker/stakeholders (D.A.R.E. America), who decided not to implement the TCYL program until further revisions and testing had been done to insure the intervention achieved the primary goal of reducing substance use in adolescents.

Dissemination of Findings

The findings of any program evaluation should be broadly disseminated to facilitate decision-making by other program implementers, stakeholders, and researchers. The ASAPS study dissemination process included annual reports and presentations to RWJF and other stakeholders. Presentations were made at professional conferences and a large number of peer-reviewed articles on the study have been published (in addition to those articles cited in this chapter, please see Brown et al. 2008; DesJarlais et al. 2006; Hammond et al. 2008; Merrill et al. 2006; Sloboda et al. 2008; Teasdale et al. 2013; Tonkin et al. 2008).

All student survey data were also made available to the public through the *Inter-university Consortium for Political and Social Research* (ICPSR), and is available to researchers and students at: https://www.icpsr.umich.edu/icpsrweb/landing.jsp. By providing other evaluators, program implementers, and researchers access to the data and findings for this evaluation, the evaluators anticipate improvements in not only in substance abuse prevention programming, but in evaluation research as well.

References

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32,* 665–683.

Arthur, M. W., & Blitz, C. (2000). Bridging the gap between science and practice in drug abuse prevention through needs assessment and strategic community planning. *Journal of Community Psychology, 28,* 241–255.

Arthur, M. W., Hawkins, J. D., Pollard, J., Catalano, R. F., & Baglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency and other adolescent problem behaviors: The Communities that Care Youth Survey. *Evaluation Review, 26,* 575–601.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Baranowski, T., & Stables, G. (2000). Process evaluation of the 5-a-day projects. *Health Education and Behavior, 27,* 157–166.

Belmont Report. (1979). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. Available at: http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 33,* 5–18.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives—The classification of educational goals—Handbook 1: Cognitive domain*. London, WI: Longmans, Green & Co. Ltd.

Botvin, G. J., & Griffin, K. W. (2003). Drug abuse prevention curricula in schools. In Z. Sloboda & W. J. Bukoski (Eds.), *Handbook of drug abuse prevention: Theory, science, and practice*. New York, NY: Springer.

Brown, C. H. (2006). Design principles and their application in preventive field trials. In Z. Sloboda & W. J. Bukoski (Eds.), *Handbook of drug abuse prevention: Theory, science, and practice*. New York, NY: Springer.

Brown, C. H., Wang, W., Kellam, S. G., Muthen, B. O., Petras, H., Toyinbo, P., et al. (2008). Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence, 95,* S74–S104.

Bukoski, W. J. (2003). The emerging science of drug abuse prevention. In Z. Sloboda & W. J. Bukoski (Eds.), *Handbook of drug abuse prevention: Theory, science, and practice*. New York, NY: Springer.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 15,* 195–207.

Carson, K. V., Brinn, M. P., Labiszewski, N. A., Esterman, A. J., Chang, A. B., & Smith, B. J. (2011). Community interventions for preventing smoking in young people. *Cochrane Database of Systematic Reviews*, Issue 7.

Castro, F. G., Morera, O. S., Kellison, J. G., & Aguirre, K. M. (2014). Mixed methods research design for prevention science: Methods, critiques, and recommendations. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science*. New York, NY: Springer.

Cegala, D. J. (1981). Interaction involvement: A cognitive dimension of communicative competence. *Communication Education, 30,* 109–121.

Centers for Disease Control and Prevention. (2004). Methodology of the Youth Risk Behavior Surveillance System. *Morbidity and Mortality Weekly Report*, *53* (No. RR-12).

Centers for Disease Control and Prevention. (2011). U.S. Department of Health and Human Services Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation. *Introduction to program evaluation for public health programs: A self-study guide.* Atlanta, GA: U.S. Centers for Disease Control and Prevention.

Center for Substance Abuse Prevention (CSAP). (1999). *Core measure initiative phase i recommendations, December 1999*. Washington, DC: Substance Abuse and Mental Health Services Administration.

Collins, L. M. (1994). Some design, measurement, and analysis pitfalls in drug abuse prevention research and how to avoid them: Let your model be your guide. In A. Cázares & L. A. Beatty (Eds.), *Scientific methods for prevention intervention research.* National Institute on Drug Abuse Research Monograph Series Number 139. Rockville, MD: National Institute on Drug Abuse.

Collins, L. M., & Flaherty, B. P. (2006). Methodological considerations in prevention research. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science*. New York, NY: Springer.

Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia, 4,* 3–11.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin.

DesJarlais, D. C., Sloboda, Z., Friedman, S. R., Tempalski, B., McKnight, C., & Braine, N. (2006). Diffusion of the D.A.R.E. and syringe exchange programs. *American Journal of Public Health, 96,* 1354–1358.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research: Theory & Practice, 18,* 237–256.

Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., et al. (2011). Evidence-based practice in school substance use prevention: Fidelity of implementation under real-world conditions. *Health Education Research, 26,* 361–371.

Ennett, S. T., Ringwalt, C. L., Thorne, J., Rohrbach, L. A., Vincus, A. A., Simons-Rudolph, A., et al. (2003).

A comparison of current practice in school-based substance use prevention programs with meta-analysis findings. *Prevention Science, 4,* 1–14.

Fishbein, D. H., & Ridenour, T. A. (2013). Advancing transdisciplinary translation for prevention of high-risk behaviors: Introduction to the special issue. *Prevention Science, 14,* 201–205.

Flay, B. R., Snyder, F., & Petraitis, J. (2009). The theory of triadic influence. In R. J. DiClemente, M. C. Kegler, & R. A. Crosby (Eds.), *Emerging theories in health promotion practice and research* (2nd ed.). New York, NY: Jossey-Bass.

Foxcroft D. R., & Tsertsvadze A. (2011). Universal multi-component prevention programs for alcohol misuse in young people. *Cochrane Database of Systematic Reviews*, Issue 9.

Gantt, H. L. (1974). *Work, Wages and Profit, published by the Engineering Magazine, New York, 1910; republished as Work, Wages and Profits*. Easton, PA: Hive Publishing Company.

Garbarino, J. (1978). The role of schools in socialization to adulthood. *Education Forum, 42,* 169–182.

Goldstein, A., & McGinnis, E. (1997). *Skillstreaming the adolescent: New strategies and perspectives for teaching prosocial skills*. Champaign, IL: Research Press.

Gorman, D. M., & Conde, E. (2007). Conflict of interest in the evaluation and dissemination of "model" school-based drug and violence prevention programs. *Evaluation and Program Planning, 30,* 422–429.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576.

Hallfors, D., & Godette, D. (2002). Will the 'principles of effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Education Research, 17,* 461–470.

Hammond, A., Sloboda, Z., Tonkin, P., Stephens, R. C., Teasdale, B., Grey, S. F., et al. (2008). Do adolescents perceive police officers as credible instructors of substance abuse prevention programs?. *Health Education Research, 23,* 682–696.

Hatry, H. P., Wholey, J. S., & Newcomer, K. E. (2010). Evaluation challenges, issues, and trends. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*. San Francisco, CA: Jossey Bass.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112,* 64–105.

Jackson, C., Gesses, R., Haw, S., & Frank, J. (2012). Interventions to prevent substance use and risky sexual behavior in young people: A systematic review. *Addiction, 107,* 733–747.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2013). *Monitoring the future national survey results on drug use, 1975–2012: Volume I, secondary school students*. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Kumar, R., O'Malley, P. M., Johnston, L. D., & Laetz, V. B. (2013). Alcohol, tobacco, and other drug use prevention programs in U.S. schools: A descriptive summary. *Prevention Science, 14,* 581–592.

Kumpfer, K. L., & Alvarado, R. (2003). Family-strengthening approaches for the prevention of youth problem behaviors. *American Psychologist, 58*(6-7), 457–465.

McKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614.

McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*. San Francisco, CA: Jossey Bass.

Merrill, J. C., Pinsky, L., Killeya-Jones, L. A., Sloboda, Z., & Dilascio, T. (2006). Substance abuse prevention infrastructure: A survey-based study of the organizational structure and function of the D.A.R.E program. *Substance Abuse Treatment, Prevention, and Policy, 6,* 1–25.

Mohr, L. B. (1995). *Impact Analysis for program evaluation*. Thousand Oaks, CA: SAGE.

National Institute on Drug Abuse. (2003). *Preventing drug use among children and adolescents: A research-based guide.* NIH Publication No. 04-4212 (A). Bethesda, MD: National Institute on Drug Abuse.

O'Connor, T. G., & Rutter, M. (1996). Risk mechanisms in development: Some conceptual and methodological considerations. *Developmental Psychology, 32,* 787–795.

Pandiani, J. A., Banks, S. M., & Schacht, L. M. (1998). Personal privacy versus public accountability: A technological solution to an ethical dilemma. *The Journal of Behavioral Health Services & Research, 25,* 456–463.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research, 34,* 1189–1208.

Pedhazur, E. J., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. New York, NY: Psychology Press, Taylor & Francis.

Petras, H., & Sloboda, Z. (2014). A conceptual foundation for prevention. In Z. Sloboda & H. Petras (Eds.), *Advances in prevention science. Volume 1: Defining prevention science*. New York, NY: Springer.

Piaget, J. (1973). *Main trends in psychology*. London, UK: George Allen & Unwin.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raymond, M. R. (1986). Missing data in evaluation research. *Evaluation and the Health Professions, 9,* 395–420.

Rodi, M. S., & Paget, K. D. (2007). Where local and national evaluators meet: Unintended threats to ethical evaluation practice. *Evaluation and Program Planning, 30,* 416–421.

Schinke, S. P., Botvin, G. J., & Orlandi, M. A. (1991). Substance abuse in children and adolescents. In S. P. Schinke, G. J. Botvin, & M. A. Orlandi (Eds.), *Substance abuse in children and adolescents: Evaluation and intervention.* Newbury Park, CA: SAGE.

Schwandt, T. A. (2007). Expanding the conversation on evaluation ethics. *Evaluation and Program Planning, 30,* 400–403.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Simons, H. (2006). Ethics in evaluation. In I. Shaw, I. Graham, R. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation*. Thousand Oaks, CA: SAGE Publications Inc.

Sloboda, Z. (2009). School prevention. In C. Leukefeld, T. Gullotta, & M. S. Tindall (Eds.), *Handbook on adolescent substance abuse prevention and treatment: Evidence-based practices*. New York, NY: Springer Academic Publishing.

Sloboda, Z. (2015a). "Read my lips"—Empty words: The semantics of institutionalized flawing. *Substance Use and Misuse, 16,* 1–6.

Sloboda, Z. (2015b). Vulnerability and risks: Implications for understanding etiology and drug use prevention. In L. M. Scheier (Ed.), *Handbook of adolescent drug use prevention: Research, intervention strategies, and practice*. Washington, DC: American Psychological Association.

Sloboda, Z., Pyakuryal, A., Stephens, P., Teasdale, B., Forrest, D., Stephens, R. C., et al. (2008). Reports of substance abuse programming available in schools. *Prevention Science, 9,* 276–287.

Sloboda, Z., Stephens, P., Pyakuryal, A., Teasdale, B., Stephens, R. C., Hawthorne, R. D., et al. (2009a). Implementation fidelity: The experience of the adolescent substance abuse prevention study. *Health Education Research, 24,* 394–406.

Sloboda, Z., Stephens, R. C., Stephens, P. C., Grey, S. F., Teasdale, B., Hawthorne, R. D., et al. (2009b). The adolescent substance abuse prevention study: A randomized field trial of a universal substance abuse prevention program. *Drug and Alcohol Dependence, 102,* 1–10.

Stephens, P. C., Sloboda, Z., Stephens, R. C., Marquette, J. F., Hawthorne, R. D., & Williams, J. (2009). Universal school-based substance abuse prevention programs: Modeling targeted mediators and outcomes for adolescent cigarette, alcohol and marijuana use. *Drug and Alcohol Dependence, 102,* 19–29.

Stephens, R. C., Thibodeaux, L., Sloboda, Z., & Tonkin, P. (2007). Research note: An empirical study of adolescent student attrition. *Journal of Drug Issues, 37,* 475–488.

Teasdale, B., Stephens, P. C., Sloboda, Z., Grey, S. F., & Stephens, R. C. (2009). The influence of program mediators on outcomes for substance users and non-users at baseline. *Drug and Alcohol Dependence, 102,* 11–18.

Teasdale, B., Stephens, P. C., Sloboda, Z., Stephens, R. C., & Grey, S. F. (2013). The effect of Hurricane Katrina on adolescent feelings of social isolation. *Social Science Quarterly, 94,* 490–505.

The Robert Wood Johnson Foundation. RWJ website: http://www.rwjf.org/en/library/research/2009/06/the-adolescent-substance-abuse-prevention-study.html

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27,* 237–246.

Tonkin, P., Sloboda, Z., Stephens, R. C., Teasdale, B., & Grey, S. F. (2008). Is the receptivity of substance abuse prevention programming impacted by students' perceptions of the instructor? *Health Education and Behavior, 36,* 724–745.

Torres, R. T. (1991). Improving the quality of internal evaluation: The evaluator as consultant-mediator. *Evaluation and Program Planning, 14,* 189–198.

United Nations Office on Drug Use and Crime (UNODC). (2013). International standards for drug use prevention. http://www.unodc.org/unodc/en/prevention/prevention-standards.html

U.S. Department of Health and Human Services Centers for Disease Control and prevention. (2011). Office of the Director, Office of Strategy and Innovation. *Introduction to program evaluation for public health programs: A self-study guide*. Atlanta, GA: Centers for Disease Control and Prevention.

W. K. Kellogg Foundation. (2004a). *Logic model development guide*. W. K. Kellogg Foundation, Battle Creek, Michigan. Accessed June 2, 2014 at: http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

W. K. Kellogg Foundation (2004b). *W. K. Kellogg Foundation evaluation handbook*. Battle Creek, MI. downloaded 5/21/2014 at: http://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). New Jersey: Prentiss Hall.

Index

Note: Page numbers followed by *f* and *t* indicate figures and tables respectively

© Springer International Publishing AG 2017
J.B. VanGeest et al. (eds.), *Research Methods in the Study of Substance Abuse*, DOI 10.1007/978-3-319-55980-3